Journal of Teaching in Physical Education

The Influence of Content Knowledge on Pedagogical Content Knowledge: An Evidence-Based Practice for Physical Education

Journal:	*Journal of Teaching in Physical Education*
Manuscript ID	JTPE.2017-0168.R2
Manuscript Type:	Article
Keywords:	Intervention studies, Professional knowledge, physical education

Running head: PEDAGOGICAL CONTENT KNOWLEDGE

The Influence of Content Knowledge on Pedagogical Content Knowledge: An Evidence-Based Practice for Physical Education

Abstract

Purpose: We conducted a retroactive analysis of teacher and student data from two randomized group trials and one well-controlled quasi-experimental group trial focused on improving pedagogical content knowledge (PCK) and student performance.

Method: Seven teachers and 32 classes were investigated. PCK was measured using four variables: task selection, representation, adaption, and an aggregate variable called total PCK. Student data are reported as percentages of correct performance. Data are reported descriptively using effect sizes (ES).

Results: The studies generated 35 ES across four teachers and one student performance variable. All ES exceeded the U.S. Department of Education's What Works Clearinghouse .25 standard deviation criterion for a "substantively important" effect and all ES exceeded Cohen's criteria of .8 for a large effect.

Discussion: Findings from this study support a focus on professional development of teachers' content knowledge as an evidenced-based practice for improving the PCK of teachers and in turn student performance.

KEY WORDS, Intervention studies, Professional knowledge, Physical Education

Unpacking Content in Pedagogical Content Knowledge: A Synthesis of Three Experimental Studies in Physical Education

It was the famous educator, Madeline Hunter who first made popular the idea that "*teaching is decision making"* (Hunter, 1979). Specifically, she argued that teachers' decision-making included applying professional and subject specific knowledge and judgment in their work (Kennedy, 2016). Moreover, she believed that the decisions teachers made could be placed into one of three categories: (a) what content to teach, (b) what students will do to learn the content, and (c) what pedagogies teachers will employ to facilitate student learning. These categories are as relevant today as when Hunter proposed them. Contemporary scholars, however, emphasize that the decisions made within each of these categories are also influenced by socio-cultural contexts, educational standards, and school values (Kirk, 2010).

Professional knowledge and subject specific knowledge underpinning the decisions a teacher makes are grounded in the assumption that they work (What Works Clearinghouse, 2014). Those responsible for teacher preparation and professional development should, in turn, focus on providing prospective and in-service teachers with knowledge that has already been proven to be useful (Darling-Hammond & Bransford, 2007; Kennedy, 2016; Ward, 2016). Mirroring the larger educational community, there have been calls for sport pedagogists to develop this kind of evidence-based practice for physical educators to employ (Hastie 2016; Institute of Medicine, 2013; McKenzie & Lounsbery, 2013; Ward, 2013). To date, however, the field has not responded to these calls or even developed criteria to determine what counts as evidence-based practice. The What Works Clearinghouse (2014) defined empirically tested educational interventions to be "substantively important" (i.e., evidence-based) when an effect size of .25 of a standard deviation or larger was achieved, even though statistical significance may not have been reached.

At least one issue that arises from the use of research in creating evidence-based practice is whether the research will work with teachers. Green (2008) has argued that an evidence-based practice should be based on practice-based evidence. His point was that there must be consideration of teachers' usage of any practice. Green (2008) suggested that teachers should be involved in the development of evidence-based practice to ensure both its usability by teachers and it sustainability as an educational practice. Such a position recognizes that teachers will always need to adapt an evidence-based practice for the varying education contexts they find themselves.

The professional judgments of teachers are not only informed by evidenced-based practice, but they are also informed by craft knowledge. Craft knowledge is "that part of professional knowledge which teachers acquire primarily through their practical experience in the classroom" (Brown & McIntyre, 1993, p. 17). We argue that evidence-based practice is not a panacea nor a substitute for a teacher's professional judgments, but nor is evidence-based practice something that can be ignored by teachers who are professionals. It is clear that teachers play an important role in operationalizing evidence-based practice that will inform their decision making.

Nowhere is the role of decision making more evident in teaching, than in the use of pedagogical content knowledge (PCK; Shulman, 1987). PCK represents the epitome of the application of professional knowledge and the use of professional judgment. In the next section, we describe pertinent theoretical and empirical literature concerned with PCK that underpinned the research reported in this article.

Theoretical and Empirical Background

Shulman (1986; 1987) was arguably the first scholar to stress the importance that content knowledge played in teaching effectiveness. He was also the first to differentiate content knowledge from PCK. Shulman (1986; 1987) argued the PCK was informed by a number of knowledge bases, and these served the function of providing the teacher with knowledge to transform the content in ways that made it understandable to learners. Since Shulman's early theoretical work, a growing body of research has examined PCK in instructional contexts (e.g., Depaepe, Verschaffel, & Kelchtermans, 2013; Ward & Ayvazo, 2016). This theoretical work has been characterized by a lack of consensus between sets of researchers, particularly concerning the knowledge bases that inform PCK (Depaepe et al., 2013). In physical education, however, most scholars have viewed PCK as being informed and influenced by teachers' knowledge of students (e.g., how students learn, develop, the cultures in which students exist), knowledge of pedagogy (e.g., effective management and instruction), knowledge of curriculum (e.g., curricular models), knowledge of context (e.g., social class and local standards for student behavior), and content (Ward & Ayvazo, 2016). Ward and Ayvazo (2016) note that the rationale for describing the influence of a particular knowledge base on PCK often varies according to the epistemological perspectives of different author.

The studies on which we reported in this article are grounded in the epistemology of radical behaviorism (Skinner, 1965) and the science of applied behavior analysis (Cooper, Heron, & Heward, 2011; Siedentop & Rushall, 1972; Ward, 2006). In behavior analysis, behavior is influenced by prior learning history and by the immediate current events. Ward, Kim, Ko, and Li (2015) defined PCK behaviorally as:

> A focal point, a locus, defined as such as an event in time (and therefore specific contextually) where teachers make decisions in terms of content based on their understandings of a number of knowledge bases (e.g., pedagogy, learning, motor development, students, contexts, and curriculum). (p. 2)

Unpacking this definition, one begins with the view that PCK is a behavior. Ward et al. (2015) conception of PCK argued that the construct can be observed in lessons as a series of decisions operationalized as instructional representations (e.g., demonstrations), instructional tasks and their progression, feedback, and modifications of instructional tasks to accommodate students' responses to tasks and feedback. Each decision a teacher makes in this regard involves applying their prior knowledge (e.g., knowledge of the student) and is influenced by the lesson's current context (e.g., equipment, space, student performances) so that the next decision is potentially informed by the previous decision. The extent to which this occurs is a function of the consequences of the teachers' behaviors. Teaching behaviors are acquired and maintained as a result of reinforcement and stopped by the absence of reinforcement and/or punishment (Cooper et al., 2007). Reinforcers vary across individuals and what is a reinforcer for one teacher may not be for another. Examples of potential reinforcers found in teaching include students being successful, staying on-task, and following directions. Examples of potential punishers would be a lack of student success, students complaining, and off task behaviors.

From a behavioral perspective, PCK can be observed and measured when teachers write lesson plans, talk about what they would do in a particular lesson or context, or when they complete tests that measure their PCK. Viewing PCK as an observed behavior is not only a behavioral perspective. Many studies of PCK though not behavioral in nature, have relied on a variety of similar data collection strategies to report PCK including discourse, written artifacts, lesson plans, and direct observation of teachers (Amade-Escot, 2005; Chen, 2004; Housner, Gomez, & Griffey, 1993; Rovegno, 1992; Rovegno, Chen, & Todorovich, 2003; Tsangaridou, 2002; Schempp, Manross, Tan, & Fincher, 1998).

Most studies in physical education conceptualize the knowledge bases that most inform PCK as knowledge of students, pedagogy, curriculum, context, and content (Ward & Ayvazo, 2016). This position is strongly influenced by the Shulman's (1987) conceptualization, the work of Grossman, Schoenfeld, and Lee (2005) and Ball, Thames, and Phelps (2008) in general education, and the work of Rovegno (1995) in physical education. Knowledge of students includes knowledge of how students learn, their developmental characteristics, characteristics of their culture, and what a teacher has learned from previous students that would indicate both their prior knowledge and the ability levels of similar students, as well as their knowledge of the students in this particular class. Knowledge of pedagogy includes not just the basic pedagogies such as class control, class organization, and the instructional techniques; it also overlaps with knowledge of curriculum such as knowing games pedagogies, cooperative learning models, or the Sport Education curriculum (Metzler, 2011). Knowledge of context includes knowledge of the resources that are available and the socio-cultural context of the community in which the school is located, district policies, and standards for learning established by the district and the state. Content knowledge as knowledge base that informs PCK, can be differentiated as two sub-domains (Ball et al., 2008; Ward, 2009). Common content knowledge (CCK) refers to knowledge of the technique and the tactics of a movement and the rules governing its performance (Ball et al., 2008; Ward, 2009). Specialized content knowledge (SCK) includes knowledge of the instructional representations of CCK, instructional tasks to teach CCK, and errors that students can make associated with those tasks (Ward, 2009). The relationship between SCK and PCK is often misunderstood.

Understanding the CCK of a movement is central to teaching it. Teachers cannot teach the forehand in tennis if they do not know the grip and the technique of the shot, nor can they teach the grapevine step in a dance if they do not know it. Knowing CCK is a prerequisite to knowing SCK. Depth of SCK influences PCK in significant ways. Shulman (1987) functionally defined PCK as how content is "organized, represented, and adapted to the diverse interests and abilities of learners and presented for instruction" (p. 8). A teacher's ability to adapt instruction depends to a very large extent on the depth of their SCK. Depth of SCK can been seen in PCK of a teacher, in terms of their representations of CCK, the instructional tasks to teach CCK and the feedback that teachers use to adapt CCK to students with different performance levels. For example, a teacher may have planned to teach a skill using a series of instructional tasks. After observing students struggling with one of the tasks, the teacher may decide to add additional tasks to make it more understandable and easier to perform. Doing so requires that first the teacher has a knowledge of alternative tasks and that knowledge is derived principally from SCK. Studies have shown that many teachers lack deep understanding of SCK and this limits their ability to adapt instruction (Ingersoll, Lux, & Jenkins, 2014; Kim, 2015; Kim, Lee, Ward, & Li, 2015; Schempp et al., 1998; Ward et al., 2015).

Until recently, it has not been possible to demonstrate that changes in teacher's content knowledge directly affected their PCK or that changes in PCK directly influenced student learning (Ward & Ayvazo, 2016). The majority of existing studies in PCK in physical education have been qualitative in nature using case studies (e.g., Rovegno, 1995), expert-novice comparisons (e.g., Schempp et al., 1998), and studies of preservice teachers (e.g., Ingersoll et al., 2014). These types of studies have been foundational in describing the context and the nature of PCK in physical education, but they have not been designed to assess the relationships among content knowledge and PCK, and PCK and student performance.

This study was a retroactive analysis of data from three previously published studies using new variables and original analyses. Our purpose was to present and synthesize findings that could lead to evidence-based practice for physical education teachers in terms of how they develop PCK. The specific research questions we examined were: (a) In what ways did teachers' PCK differ before and after CCK and SCK training?; (b) In what ways did student performance differ before and after CCK and SCK training?; and (c) What were the effect sizes for teacher and student variables across studies?

Method

All three studies were part of an intentional replication effort designed to systematically replicate the independent variable, CCK and SCK training, with teachers of different backgrounds and their students. All studies were subject to the institutional review board oversight at the institutions, where the studies were conducted and informed consent was obtained for all teacher and student participants. This section is organized as follows. First, the research design and description of the studies are discussed. Second, the CCK and SCK training is described including treatment integrity. Next, teacher and student dependent variables are defined, data collection is described, and the reliability of observations are reported. Finally, the types of data analyses we used to examine and synthesize the data are reported. The analyses conducted in this study and subsequent findings are original.

Research Design and Demographic Information of Three Studies

Each of the three field-based intervention studies each had the following design characteristics: (a) the teachers taught several classes of badminton the way they would normally do so (pre-treatment) after which they received a professional development workshop on the CCK and SCK of badminton (treatment) and subsequently taught several additional classes of badminton (post-treatment); (b) the same measures of teacher behavior and student performance were examined in each study; (c) the studies were conducted in secondary school physical education settings; (d) the independent variable (CK workshop) was standardized and teacher training procedures met the same criteria; and (e) the number of pre- and post-treatment lessons taught were similar.

Study 1: Ward, Kim, Ko, and Li. (2015). This study was conducted with four middle school physical education teachers who considered badminton as their non-expert content area but who had taught badminton to grades six through eight ranging from 4-20 years. A total of 96 students were selected from the teachers' identified stratified skill-level groups (low, average and high skilled) with equal numbers of male and female students in each group. Each teacher taught two classes of students each day for six days before and after the intervention. The classes represented a convenience sample and the design was a matched quasi-experimental group trial. The study was conducted in Ohio and in North Carolina in the United States. Teacher data were collected using the video recordings of the lessons, whereas student data were collected live on every trial made by six students in each class.

Study 2: Sinelnikov, Kim, Ward, Curtner-Smith, and Li. (2016). This study was conducted with two first year physical education teachers who had little to no experience in teaching badminton to grades six to eight. A total of 48 students were selected from the teachers' identified stratified skill-level groups (low, average, and high skilled) with equal numbers of male and female students in each group. Each teacher taught two classes of students each day for six days before and after the intervention. The classes were randomly assigned to each pre-treatment or post-treatment and the design was a randomized group trial. The study was conducted in Alabama in the United States. Teacher data were collected through the use of videotaped lessons, whereas student data were collected live in each class.

Study 3: Iserbyt, Ward, and Li. (2015). This study was conducted with one high school physical education teacher who had taught badminton to grades 9-12 for 10 years and did not consider badminton as a strong content area. A total of 64 students were selected from the teachers' identified stratified skill-level groups (low, average and high skilled) with equal numbers of male and female students in each group. Each teacher taught two classes of students each day for six days before and after the treatment. The classes were randomly assigned and the design was a randomized group trial. The study was conducted in Flanders, Belgium. Both teacher and student data were collected from the video recordings of the lessons.

Treatment Description

In each study, the treatment was described as a professional development workshop intended to develop the teachers' CCK, but predominantly the SCK of badminton for secondary settings. An initial study conducted by Kim (2011) was used to standardize training and materials. The materials included a content knowledge packet consisting of descriptions of the CCK and SCK for badminton, videos of correct and incorrect performances, coding instructions and instruments, as well as observer training instructions and materials. The lead author trained all co-investigators who then trained their respective data collectors. First, individual teachers reviewed the content of the content knowledge packet prior to the workshop. The content of this packet was derived from Play Practice (Launder, 2001) and Badminton Steps to Success (Grice, 1996) for teaching five badminton skills (serve, overhead stroke, underhand stroke, smash, and drop) and basic singles/doubles strategies. Next, teachers were trained for four hours typically across two days in a workshop that consisted of three components: (a) an overview and introduction of Play Practice (Launder, 2001), (b) observation of training videos that included examples of possible errors, error corrections, task representations, task progressions, and task modifications, and (c) evaluation of the participants' understanding of the knowledge presented during the workshop through a series of questions. Teachers had to answer a total of 64 questions during the course of the workshop. When a teacher's correct answers met a criterion of 95% for each element assessed (e.g., possible errors, error corrections, task representations, task progressions, and task modifications), they were judged sufficiently well trained.

Treatment Integrity for the Workshop and Fidelity of Teacher Implementation of Lessons

To ensure that each study used the same procedures to train the teachers (i.e., treatment integrity), a checklist was used to determine if training procedures described in the treatment description above were followed. All studies reported 100% compliance with the checklist. The treatment integrity was aided by standardization through the use of the same materials, and videos of correct and incorrect performances for training.

In addition, fidelity of implementation of the tasks that were presented in the workshop were examined from an analysis of the lessons taught post-treatment (workshop). All instructional tasks were assessed using four levels of congruency with tasks presented at the workshop. The levels of task congruency were as follows: (a) used as taught in the workshop; (b) partially correct, but consistent with workshop; (c) different, but appropriate, and consistent with workshop; (d) different, but inappropriate, and not consistent with workshop. Each study reported mean fidelity of implementation for their teachers using the tasks as taught in the workshop or partially taught in the workshop. Fidelity of implementation for the Ward et al. (2015) study was 88%, the Sinelikov et al. (2016) study 77%, and the Iserbyt et al. (2015) study 91%.

Teacher Variables

Teacher variables were chosen to represent PCK. First, *task selections* by a teacher was measured whenever the teacher presented instructional tasks to the class as a whole during the lessons. Each task was judged according to two criteria: developmentally appropriate and principally appropriate. Developmental appropriateness was defined as an instructional task that was matched to the ability or readiness of the students (e.g., equipment, space, and complexity) (Kim, 2011). Principally appropriate was defined as an instructional task that was an appropriate progression relative to the previous tasks either in terms of technical development or game development (Kim, 2011).

Second, *task representations* by a teacher were categorized as either verbal or visual. Verbal representations were defined as task representations using different verbal forms during instruction and practice (e.g., correct instructions, descriptions, analogies, metaphors, cues, and feedback). Visual representations were defined as task representations using different visual forms of during instruction and practice (e.g., correct full or partial demonstrations, visual aids – task cards, diagrams, pictures, or video clips, and physical assistance).

Third, *task adaptations* by the teacher were categorized as inter- or intra-task adaptations. Inter-task adaptations were defined as task modifications between tasks for the entire class using four task categories defined by Rink (2010; i.e., informing, extending, refining, and applying tasks). Intra-task adaptations were defined as task modifications within tasks toward small groups of students or individuals. Using four task categories (modifying, refining, restating, and changing competition conditions), the number of each type of adaptation made by the teachers was measured.

We used an aggregate variable, called *total PCK* to provide an overall judgment of the PCK of the teacher. The variable was created using the above three individual teacher variables by summing the mean percentages of the three individual teacher variables and dividing by three (number of variables).

Student Variables

Every student participant performance in every lesson was coded as correct, incorrect, or other. A correct trial was defined from the literature as the demonstration of critical elements in each of primary skills in the three phases of skill performance: preparation, execution, and follow through (Grice, 1996). When students performed the critical elements correctly in two or three phases, it was coded as correct. When students performed the critical elements correctly in one or in no phases, it was coded as incorrect. When students missed hitting the shuttle due to mistakes made by themselves (e.g., not moving into position quickly enough), or if an unhittable shuttle was sent to them, or if students performed non-target movements that the teacher did not request (e.g., performing the forehand stroke when practicing the backhand stroke), it was coded as other.

Reliability of Observations. Reliability in each study was established by inter-observer agreement using the following formula: agreement divided by the sum of agreements and disagreements, then multiplied by 100 (Cooper et al., 2007). Each study reported its reliability. Ward et al. (2015) collected data on 33% of all teacher and student observations. The reliability for teacher and student data was 88% (range = 85.3-99.3% for students and range = 78.9-92.3% for teachers). Sinelnikov et al. (2016) collected data on 33% of all teacher and student observations. The reliability for teacher and student data was 89% (range = 83.8–96.6% for students and range = 81.0–91.0% for teachers). Iserbyt et al. (2015) collected data on 40% of all teacher observations and 39% of student observations. The reliability for teacher and student data was 88% (range = 84.0–90.0%).

Data Analysis

All teacher and student data were coded using event recording (Cooper, Heron, & Heward, 2007). Data analyses were conducted using the SPSS v24.0 software (SPSS, 2015). Data were analyzed in four ways. First, descriptive statistics were used for both teacher and student data. The means and standard deviations per lesson were calculated by individual teachers or their students. We used a lesson as a unit of teacher and student measurement since teacher's instructional performance varied per lesson, which impact students' success of learning. Second, using the means we calculated the mean percentages of each variable in order to show the relative standing of the individual teachers' data in relation to others in all teacher variables with the same measurement scale using the following formula: the mean divided by the highest group mean multiplied by 100. For the student data, the mean percentages of students' correct/incorrect/other variables were calculated by dividing the total number of each variable by the total trial number and then multiplying by 100. Third, we calculated effect sizes (ES) for both teacher and student variables using the means and standard deviations of two groups (pre-and post-treatment). Fourth, we computed the mean ES as recommended by Borenstein, Hedges, Higgins, and Rothstein (2009) to avoid running the risk of using a regular meta-analysis that could create measurement errors because of our small sample. Because the studies included in this analysis were similarly precise, used the same procedures, and measured the same dependent and independent variables, and because the number of the studies was small and the samples within the study were similar, reporting the mean ES creates an unweighted fixed effects meta-analysis model (Borenstein et al., 2007). The fixed-effects model assumes homogeneity of effects across the studies being combined. As such the common ES represents a summary variable that can be used to summarize the effects of training teachers in CCK and SCK and effects of training student performance as evidence-based practice.

For both the ES and meta-analysis of common ES, we used two criteria for interpreting the meta-analysis. First, Cohen's (1988) recommendations for interpreting ES was used and defines a "small" ES as .20, a "medium" ES as .50, and a "large" ES as .80. Cohen (1988) warned, however, these rules of thumb may vary within different fields of study. Second, to contextualize findings in education, we also used the U.S. Department of Education's What Works Clearinghouse (2014) ES .25 criterion for a substantively important effect.

Results

In what ways did teachers' PCK differ before and after CCK and SCK training?

The descriptive analysis of teacher data indicated that the teachers used more developmentally/principally appropriate tasks, verbal/visual representations, inter-/intra-task adaptations per lesson after developing badminton CCK and SCK. The means and standard deviations of individual teacher performance data across the three studies are presented in Table 1. Overall, the means of both developmentally/principally appropriate tasks were 0.61 (SD=0.69) in the pre-treatment classes and 2.58 (SD=0.97) the post-treatment classes. The means of verbal/visual task representations were 25.96 (SD=14.76) in the pre-treatment classes and 66.85 (SD=22.68) in the post-treatment classes. For the task adaptation variable, the means of both inter- and intra-task adaptations were 3.84 (SD=1.96) in the pre-treatment classes and 10.04 (SD=3.30) in the post-treatment classes. The means for the total PCK variable were 10.13 (SD=5.32) in the pre-treatment classes and 26.45 (SD=8.19) in the post-treatment classes.

[Insert Table 1]

To show the relative standings of the individual teachers' PCK variables in relation to others, the mean percentages of task selections, task representations, task adaptations, and total PCK in the pre-treatment and post-treatment classes are presented in Figure 1-4. We ordered the teacher data from lowest to highest based on the total PCK score reported in Figure 4. Figure 1 shows a range of the mean percentages for the task selection by the seven teachers from 0%~

23.71% in the pre-treatment classes to 30.86% ~ 100% in the post-treatment classes. Figure 2 shows a range of the mean percentages for the task representation by the teachers from 11.99%~ 48.03% in the pre-treatment classes to 49.69% ~ 100% in the post-treatment classes. Figure 3 shows a range of the mean percentages for the task adaptation by the teachers from 19.56%~ 25.50% in the pre-treatment classes to 27.81% ~ 100% in the post-treatment classes. Figure 4 shows a range of the mean percentages for the total PCK variable by the teachers from 13.59%~ 44.44% in the pre-treatment classes to 49.64% ~ 100% in the post-treatment classes. Collectively, figures 1-4 show that the teaching performances in the experimental classes were better than those in the pre-treatment classes for all teacher variables.

[Insert Figures 1-4 here]

In what ways did student performance differ before and after CCK and SCK training?

The descriptive analysis of student data indicated that the students in the post-treatment classes used more correct trials and fewer incorrect trials that those in the pre-treatment classes taught by the individual teachers. The mean percentages and standard deviations of student performance data are presented in Table 2. The mean percentages of students' correct trials per lesson were 16.10% (SD = 9.65) in the pre-treatment classes and 45.80% (SD = 14.65) in the post-treatment classes. The mean percentages of students' incorrect trials per lesson were 61.07% (SD = 10.26) in the pre-treatment classes and 32.75% (SD = 10.43) in the post-treatment classes. The mean percentages of other performance per lesson were 20.72% (SD = 6.38) in the pre-treatment classes and 21.09% (SD = 7.71) in the post-treatment classes.

[Insert Table 2]

What were the effect sizes for teacher and student variables across studies?

The analysis of teacher and students' Cohen's *d* ES per variable across the three studies is presented in Table 3. For each variable, the ranges of Cohen's *d* ES were: task selection 0.93 ~ 3.58, task representation 1.06 ~ 4.09, task adaptation 0.74 ~ 3.78, total PCK 1.11 ~ 4.85 and students' correct performance 1.53 ~ 3.67. The common ES reported as Cohen's *d* for each variable were: task selection 2.32, task representation 2.48, task adaptation 2.11, total PCK 2.44 and students' correct performance 2.35.

[Insert Table 3]

Discussion

We began this article noting that there have been increasing calls for evidenced-based practice in physical education to support the work of teachers (Hastie, 2016; Institute of Medicine, 2013; McKenzie & Lounsbery, 2013; Ward, 2013). A key element of evidence-based practice is that teaching behaviors are tied functionally to student measures, without which judgments about the effectiveness of the practice can be made. Our analyses from the two randomized group trials and one well-controlled quasi- experimental group trial provide important evidence.

Pedagogical Content Knowledge

The first analysis examined the effects of professional development training on four variables we used to define PCK. To understand the connections between PCK and the variables we used in this study they are situated within Shulman's (1987) description. PCK is "organized [*task selections*], represented [*task representations*], and adapted [*task adaptations and task selections*] to the diverse interests and abilities of learners and presented for instruction." (p. 8)

Task selection is a measure of teacher decision making and represents the SCK of the teacher (Ward & Avayzo, 2016). In the three studies, the teachers use of tasks in the post-treatment classes compared to the pre-treatment classes were substantively better. The difference between the pre-and post-treatment lessons in the selection of content can be best described as tasks used in the post-treatment lessons were incrementally progressive and aligned with the goal of the lesson. Task representations indicate a teacher's understanding of the task and how to translate content into understandable information for students. A key feature in this presentation is linking current understanding to the prior knowledge of the student. The tasks represented in the post-treatment lessons can be described using Rink's (2010) classification of content development as extension and refining with some applying tasks. Because the tasks were incrementally progressive each task built upon the previous task drawing upon the knowledge that students had of the previous task to inform the current task. This was intentional and it creates a conceptually aligned understanding of the content. Whereas in the pre-treatment lessons this depth of understanding tied to prior knowledge was less evidenced. Another feature of task representation is the use of demonstrations and description analogies and metaphors. Though it is true that more is not better, the substantive increase in aligned representations in the post-treatment classes indicates that the teachers made more effort to explain and demonstrate correctly the content they were teaching than they did in the pre-treatment classes.

Task adaptations are task selections that are made to modify the initial task selection to meet the needs of the students, small groups or even for the class as a whole. Though all the variables address the notion of adaption, in the context of Shulman's (1987) definition, task adaptations to intra and inter-task development might be one of the most observable indicators of PCK, because it best represents the notion of adaptation to needs. The data from the three studies show that the teachers made substantive efforts to adapt to student performance during lessons based on their observations of the students.

The teachers in the three studies varied significantly in their experience of teaching badminton from no prior experience (Sinelnikov et al., 2016) to experienced teacher (Iserbyt et al., 2015; Ward et al., 2015), yet the professional development training assisted them all and as we showed in the results, students in their classes benefited. Because the teachers taught before and after the treatment, teacher effects can be ruled out, and thus this data represents the growth of the teachers on these four variables.

Despite the rigor of data collection there remains a disconnect in the evidence to date between the training and the implementation of the training. These studies did not directly measure teacher knowledge and so we don't know the specific extent of their CCK and SCK prior to and following the professional development training. At the time, the studies were conducted there were no validated measures of CCK or SCK for badminton, or for any other sport. The establishment of reliable and valid measures of CCK and SCK is clearly a priority direction for future research.

Student Performance

In the three studies, data were collected on every trial made by the students in the sample. Student performance is a critical aspect of the learning process and a known predictor of learning and success in physical education (Rink, 2010; Silverman, 1985; Silverman, Subramaniam, & Woods, 1998). Behaviorally, process data of student performance represents the entire motoric engagement in badminton in the lessons. From a behavioral analytic point of view, it represents learning because the motor engagement of the performance being taught demonstrated alignment between assessment and instruction (Cooper et al., 2011). However, the motor learning literature makes the case that retention measures represent a truer measure of learning, because they demonstrate the retention of the performance following instruction (MaGill, 2011). That said future studies should conduct retention measures to provide an outcome measure and provide a measure of learning.

Ward (2013) argued, “if you use low-quality tasks, students acquire something different than if you used high-quality tasks” (p. 437), making the case that you get what you teach. Thus, we argue that what students came to understand and perform as badminton post-treatment can be defined by the quality of tasks selected the quality of the representation on those tasks and the quality of subsequent adaptations. These differences are reflected student performance measures.

Effect Sizes

The studies generated 35 ESs across the four teacher variables and the student performance variable. All ES scores except one exceeded the U.S. Department of Education’s What Works Clearinghouse .25 criterion for a “substantively important” effect by a factor of at least six and all exceeded Cohen’s criteria (i.e., > .8) for a large effect. The one exception exceeded the criterion by a factor of three. Collectively, this demonstrates substantively important effects of the professional development in developing the teachers’ PCK and also substantively important effects on correct trials of their students. The common ES representing our unweighted meta-analysis ranged from 2.11 to 2.48 standard deviations. To place this in perspective, if these were standardized assessments rather than unstandardized, the results with an ES of 1.0 would increase percentile scores from 50 to 84.

The ES we reported are large. ES can be influenced by a number of factors (Slavin, 2009), chief among these is the size of the sample. Studies with larger samples may generate ESs that are more accurate, because they are more representative of the population. This is mediated somewhat by the random assignment of teachers and students to conditions. However, Slavin (2009) notes that a large well controlled quasi-experimental study may generate more accurate ESs, than a small randomly controlled study. In the data we reviewed, the three studies both individually and cumulatively have small sample sizes for teachers. This should be considered in the interpretation of the results. However, we would note that if you cut the ESs by two thirds, an unlikely outcome with a larger sample, you would still have a substantive set of ESs all of which would exceed the criteria we used for substantively important effects. A second potential reason for high ESs also related to small samples is the effect of class and teacher confounds on the results. This typically occurs when different teachers are assigned to the pre-treatment lessons compared to the post-treatment lessons. In the studies we reviewed, this was not the case. Teachers in all studies taught both pre-and post-treatment classes ruling out teacher effects and two of the studies had random assignment of classes to conditions.

Third, de Boer, Donker and van der Werf (2014) have reported that it is typical to get higher ESs when using non-standardized measures than when compared to standardized measures. In these studies, the authors did use non-standardized measures and this is a potential reason for higher ESs and as such should be considered in the interpretation of the ESs we report. Finally, perhaps the most parsimonious explanation in the tradition of Occam's Razor (Cooper et al., 2011) for the high ESs is the low level of student performance and weak content development seen in the pre-treatment lessons accompanied by the low levels of student success.

Limitations and Strengths

As we have previously noted, the analyses conducted involved clear limitations including small sample size, a lack of retention measures to measure student learning, the lack of a direct measure of teacher content knowledge, and the duration of the instructional units, though ecologically valid, in our view too short to show the potential gains that could be made with gains in teacher knowledge. Future studies should address these shortcomings.

However, we feel confident that the effects of our measures are robust for three reasons. First, the stratified selection of students by ability represents a sampling of the students in each class. However, for students in the sample, all trials made by the students were reported and analyzed. As such these represent census level data rather than a sampling of the student participants in the study. Census data reflecting a complete picture of the use of these variables is also true for our teacher data. This is important because for the teachers it captures their PCK as we have defined it. Second, the treatment integrity representing the degree to which the professional development workshop was delivered as described in the studies was high, strengthening confidence in the internal validity of the studies. Third, the mean fidelity of implementation across the three studies by teachers was 85% lending confidence that the training from the workshop was transferred to teaching practice, and again strengthening confidence in the internal validity of the studies.

Implications for Policy

The current findings show that improving the content knowledge of teachers and in particular the SCK of teachers leads to more effective teaching and learning. The national standards for beginning teachers (SHAPE America, 2017) now require both CCK and SCK to be taught and evaluated in teacher preparation programs and our finding reinforces the rationale for this policy. Given that the duration of training for the three studies ranged from 4-6 hours the data also suggest that incorporating SCK training into continuing professional development for teachers is likely to be a small cost in terms of time to train and cost to train, because such training could be included in existing one-or two-day professional development efforts in the U.S. that often occur for a day in duration.

There have been some criticisms in the literature that these short duration professional development efforts are not effective (Gutsky & Yoon, 2009). The results presented in the current study, along with those presented by Gutsky and Yoon (2009) and Kennedy (2016) show that the time and intensity spent in professional development activities are unrelated to improvements in student outcomes. That said, we know little about the effects on student learning of sustained teacher conversations about their teaching of content beyond the training reported in studies such as we reviewed. Such studies are important directions for future research.

Conclusion

Our purpose in this study was to present and synthesize findings that could contribute to evidence-based practice for preservice and continuing professional development of teacher's content knowledge that improves PCK, and in turn, student performance. To improve student outcomes at scale, policies and practices for teachers should be based on the best evidence available. For student outcomes in secondary physical education, the evidence summarized in this study supports a focus on professional development specific to strategies used to advance CCK and SCK in teachers. Our findings support recent policy SHAPE America standards for beginning teachers

References

Amade-Escot, C. (2005). The critical didactics incidents as a qualitative method of research to analyze the content taught. *Journal of Teaching in Physical Education, 24,* 127-148.

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

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons, Ltd.

Brown, S., & McIntyre, D. (1993). Making sense of teaching. Buckingham, Open University Press.

Chen, A. (2004). Learning the skill theme approach: Salient and problematic aspects of pedagogical content knowledge. *Education, 124,* 194-212.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do.* San Francisco, CA: Jossey-Bass.

de Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, *84*, 509–545.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12–25.

Green, L. (2008). Making research relevant: If it is an evidence-based practice, where's the practice based evidence? *Family Practice, 25,* 120-124.

Grice, T. (1996). *Badminton: Steps to success* (2nd ed.). Champaign, IL: Human Kinetics.

Grossman, P. L., Schoenfeld, A., & Lee, C. D. (2005). Teaching subject matter. In L. Darling-Hammond, J. Bransford, P. LePage, K. Hammerness, & H. Duffy (Eds.). *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 201-231). San Francisco, CA: Jossey Bass.

Gutsky, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*, 495-500.

Hastie, P. (2016). The philosophy of physical education: A new perspective. *Journal of the Philosophy of Sport, 43*, 1-3.

Housner, L. D., Gomez, R. L., & Griffey, D. C. (1993). Pedagogical knowledge structures in prospective teachers: relationships to performance in a teaching methodology course. *Research Quarterly for Exercise and Sport, 64*, 167-177.

Hunter, M. (1979). Teaching is decision making. *Educational Leadership*, *37*, 62-67.

Ingersoll, C., Lux, G.K., & Jenkins, J. M. (2014). There is no lack of people to go to for support: Policy, community, and individual influences on one pre-service teacher's knowledge development. *Global Journal of Health and Physical Education Pedagogy, 3*, 37-53.

Institute of Medicine. (2013). *Educating the student body: Taking physical activity and physical education to school.* Washington, DC: The National Academies.

Iserbyt, P., Ward, P., & Li, W. (2015). Effects of improved content knowledge on pedagogical content knowledge and student performance in physical education. *Physical Education and Sport Pedagogy, 22*, 71-88.

Kennedy, M. (2016). How does professional development improve teaching? *Review of Educational Research, 86,* 945– 980.

Kim, I. (2011). *The effects of a badminton content knowledge workshop on middle school physical education teachers' pedagogical content knowledge and student learning.* Unpublished doctoral dissertation. The Ohio State University, Columbus, OH.

Kim, I. (2015). Exploring changes to a teachers' teaching practices and student learning through a volleyball content knowledge workshop. *European Physical Education Review*, *22*, 1-18 DOI: 10.1177/1356336X15599009

Kim, I., Lee, Y. S. Ward, P., & Li, W. (2015). A critical examination of content knowledge courses in physical education teacher education programs. *Journal of Teaching in Physical Education,34*, 59-75.

Kirk. D. (2010). *Physical education futures.* Milton, UK: Routledge.

Launder, A. G. (2001). *Play practice: The games approach to teaching and coaching sports*. Champaign, IL: Human Kinetics.

MaGill, R.A. (2011). *Motor learning and control: Concepts and applications*. New York, NY: McGraw-Hill.

McKenzie, T. L., & Lounsbery, M. A. F. (2013). Physical education teacher effectiveness in a public health context. *Research Quarterly for Exercise and Sport, 84,* 419–430.

Metzler, M. (2011). *Instructional models for physical education* (3rd ed.). Scottsdale, AZ: Holcomb Hathway.

Rink, J. (2010). *Teaching physical education for learning* (6th ed.). Boston, MA: McGraw-Hill Higher Education.

Rovegno, I. (1992). Learning to teach in a field-based method course: The development of pedagogical content knowledge. *Teaching and Teacher Education, 8,* 69-82.

Rovegno, I. (1995). Theoretical perspectives on knowledge and learning and a student teacher's pedagogical content knowledge of dividing and sequencing subject matter. *Journal of Teaching in Physical Education, 14, 28*4-304.

Rovegno, I., Chen, W., & Todorovich, J. (2003). Accomplished teachers' pedagogical content knowledge of teaching dribbling to third grade children. *Journal of Teaching in Physical Education, 22,* 426-449.

Schempp, P. G., Manross, D., Tan, S. K. S., & Fincher, M. D. (1998). Subject expertise and teachers' knowledge. *Journal of Teaching in Physical Education, 17,* 342-356.

SHAPE America. (2017). *National standards for initial physical education teacher education.* Retrieved July 10, 2017, from http://www.shapeamerica.org/accreditation/upload/2017-SHAPE-America-Initial-PETE-Standards-and-Components.pdf

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

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57,* 1-22.

Siedentop, D., & Rushall, B. (1972). An operant model for skill acquisition. *Quest, 17,* 82–90.

Silverman, S. (1985). Relationship of engagement and practice trials to student achievement. *Journal of Teaching in Physical Education*, *5,* 13-21.

Silverman, S., Subramaniam, P.R., & Woods, A.M. (1998). Task structures, student practice, and student skill level in physical education. *Journal of Educational Research*, *91*, 298-306.

Sinelnikov, O., Kim, I., Ward, P., Curtner-Smith, M., & Li, W. (2016). Changing beginning teachers' content knowledge and its effect on student learning. *Physical Education and Sport Pedagogy*, *4,* 420-440.

Skinner, B. F. (1965). The technology of teaching. *Proceedings of the Royal Society, B. 162,* 427-443.

Slavin, R. E. (2009). *Educational psychology: Theory and practices* (9th ed.). Upper Saddle River, NJ: Pearson/Merrill.

SPSS Institute. (2015). *IBM SPSS statistics for windows* (Version 24.0). Armonk, NY: IBM Corp.

Tsangaridou, N. (2002). Enacted pedagogical content knowledge in physical education: A case study of a prospective classroom teacher. *European Physical Education, 8*, 21-36.

Ward, P. (2006). The philosophy, science and application of behavior analysis in physical education. In D. Kirk., D. Macdonald, & M. O'Sullivan, *The Handbook of Physical Education* (pp. 3-21). London, UK: Sage Publications.

Ward, P. (2009). Content matters: Knowledge that alters teaching. In L. Housner, M. Metzler, P. Schempp, & T. Templin, *Historic Traditions and Future Directions of Research on Teaching and Teacher Education in Physical Education* (pp. 345-356). Morgantown, WV: Fitness Information Technology.

Ward, P. (2013). The role of content knowledge in conceptions of teaching effectiveness in physical education. *Research Quarterly for Exercise and Sport, 84,* 431–440.

Ward, P. (2016). Policies, agendas and practice influencing doctoral education in physical education teacher education, *Quest*, *68*, 420-439.

Ward, P., & Ayvazo, S. (2016). Pedagogical content knowledge: Conceptions and findings in physical education. *Journal of Teaching in Physical Education, 35,* 194-207.

Ward, P., Devent, F., Lee, Y. S., Ko, B., Kim, I., & Tao, W. (2017). Using content maps to measure content development in physical education: Validation and application. *Journal of Teaching in Physical Education, 36*, *20-31.*

Ward, P., Kim, I., Ko, B., & Li, W. (2015). Effects of improving teachers' content knowledge on teaching and student learning in physical education. *Research Quarterly for Exercise and Sport*, *86*, 130-139.

What Works Clearinghouse. (2014). *Procedures and standards handbook* (V. 3). Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

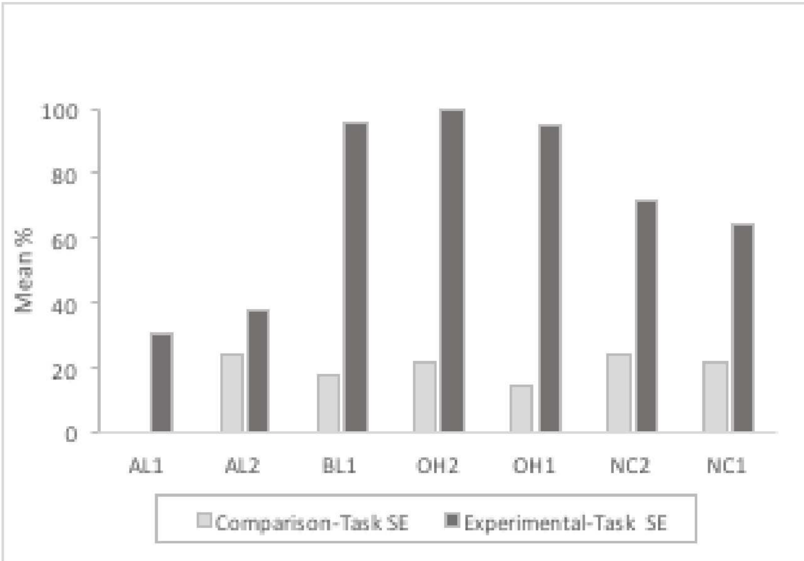

Figure 1. Mean percentages of teachers' task selections in the comparison and experimental classes

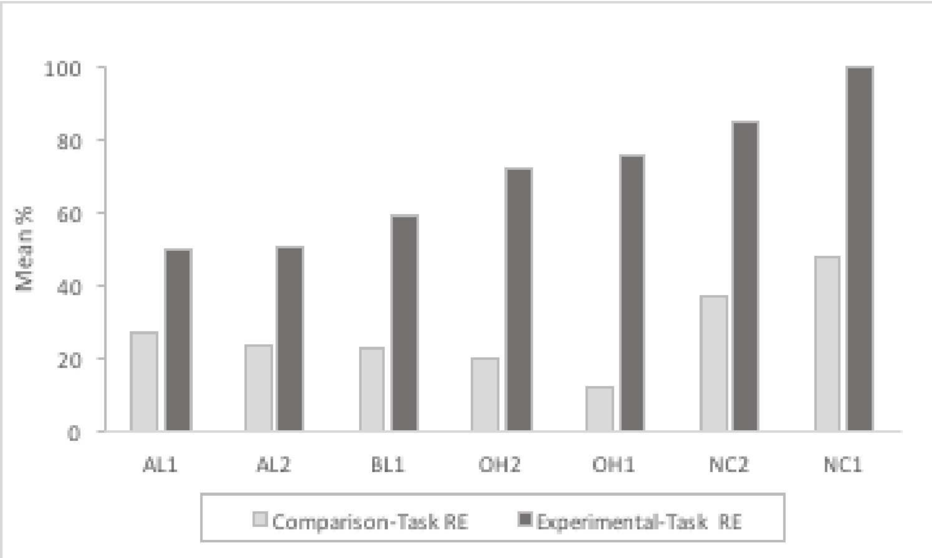

Figure 2. Mean percentages of teachers' task representations in the comparison and experimental classes

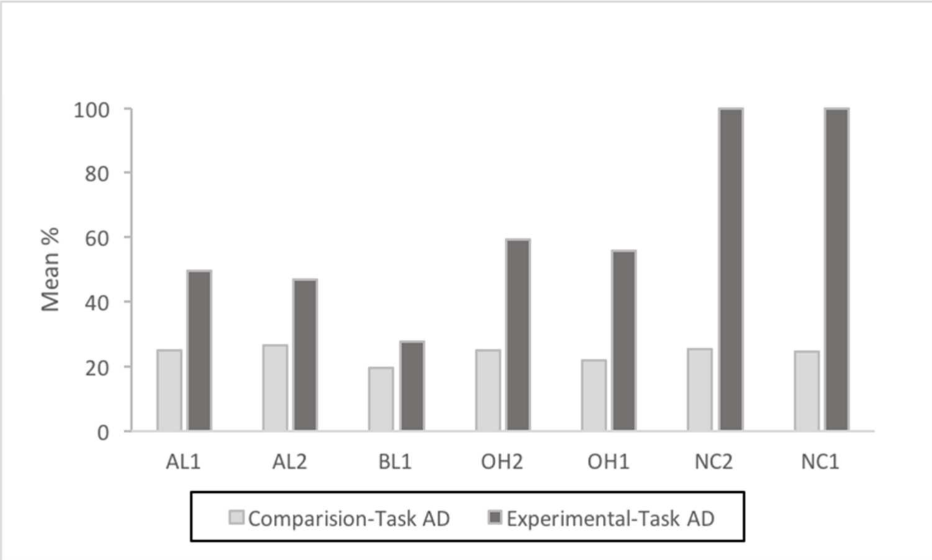

Figure 3. Mean percentages of teachers' task adaptations in the comparison and experimental classes

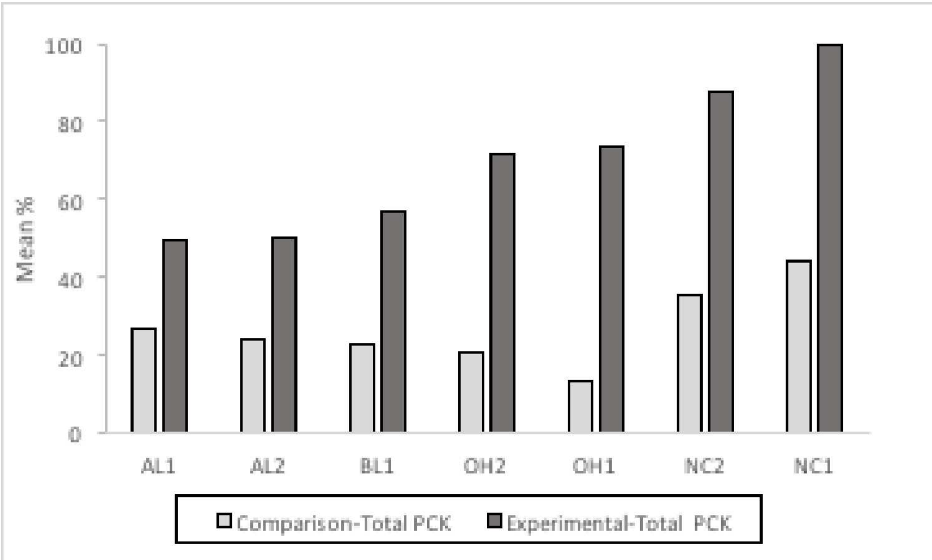

Figure 4. Mean percentages of teacher's total PCK in the comparison and experimental classes

Table 1. Descriptive statistics of teacher variables by treatment condition

Teachers	Treatment (N)		Task Selection	Task Representation	Task Adaptation	Total PCK
AL 1	Control	*M*	0.00	26.00	4.00	10.00
	(N = 12)	*SD*	0.00	15.90	4.25	6.08
	Experimental	*M*	1.83	47.33	7.92	18.77
	(N = 12)	*SD*	1.00	23.59	4.34	9.39
AL 2	Control	*M*	0.83	22.42	4.25	9.17
	(N = 12)	*SD*	0.58	12.20	2.14	4.58
	Experimental	*M*	1.33	47.92	7.50	18.92
	(N = 12)	*SD*	0.49	23.38	2.11	8.25
BL 1	Control	*M*	0.61	21.90	3.13	8.55
	(N = 24)	*SD*	0.50	22.32	1.80	7.81
	Experimental	*M*	3.34	56.60	4.45	21.46
	(N = 24)	*SD*	1.24	18.26	1.78	6.75
OH 2	Control	*M*	0.75	18.75	4.02	7.83
	(N = 12)	*SD*	0.87	8.53	2.13	3.46
	Experimental	*M*	3.50	68.60	9.50	27.20
	(N = 12)	*SD*	0.71	35.47	2.07	12.13
OH 1	Control	*M*	0.50	11.42	3.50	5.14
	(N = 12)	*SD*	0.80	5.55	1.09	1.88
	Experimental	*M*	3.33	71.58	8.92	27.94
	(N = 12)	*SD*	0.78	20.04	3.32	6.37
NC 2	Control	*M*	0.83	45.75	4.08	13.44
	(N = 12)	*SD*	1.11	24.18	2.23	5.35
	Experimental	*M*	2.50	80.67	16.00	33.06
	(N = 12)	*SD*	1.68	22.49	5.34	8.05
NC 1	Control	*M*	0.75	45.75	3.92	16.81
	(N = 12)	*SD*	0.97	24.18	1.83	8.10
	Experimental	*M*	2.25	95.25	16.00	37.83
	(N = 12)	*SD*	0.87	15.73	4.13	6.40
Total	Control	*M*	0.61	25.96	3.84	10.13
	(N = 96)	*SD*	0.69	14.76	1.96	5.32
	Experimental	*M*	2.58	66.85	10.04	26.45
	(N = 96)	*SD*	0.97	22.68	3.30	8.19

Table 2. Descriptive statistics of teacher variables by treatment condition

Teachers	Treatment (N)		Task Selection	Task Representation	Task Adaptation	Total PCK
AL 1	Control	*M*	0.00	26.00	4.00	10.00
	(N = 12)	*SD*	0.00	15.90	4.25	6.08
	Experimental	*M*	1.83	47.33	7.92	18.77
	(N = 12)	*SD*	1.00	23.59	4.34	9.39
AL 2	Control	*M*	0.83	22.42	4.25	9.17
	(N = 12)	*SD*	0.58	12.20	2.14	4.58
	Experimental	*M*	1.33	47.92	7.50	18.92
	(N = 12)	*SD*	0.49	23.38	2.11	8.25
BL 1	Control	*M*	0.61	21.90	3.13	8.55
	(N = 24)	*SD*	0.50	22.32	1.80	7.81
	Experimental	*M*	3.34	56.60	4.45	21.46
	(N = 24)	*SD*	1.24	18.26	1.78	6.75
OH 2	Control	*M*	0.75	18.75	4.02	7.83
	(N = 12)	*SD*	0.87	8.53	2.13	3.46
	Experimental	*M*	3.50	68.60	9.50	27.20
	(N = 12)	*SD*	0.71	35.47	2.07	12.13
OH 1	Control	*M*	0.50	11.42	3.50	5.14
	(N = 12)	*SD*	0.80	5.55	1.09	1.88
	Experimental	*M*	3.33	71.58	8.92	27.94
	(N = 12)	*SD*	0.78	20.04	3.32	6.37
NC 2	Control	*M*	0.83	45.75	4.08	13.44
	(N = 12)	*SD*	1.11	24.18	2.23	5.35
	Experimental	*M*	2.50	80.67	16.00	33.06
	(N = 12)	*SD*	1.68	22.49	5.34	8.05
NC 1	Control	*M*	0.75	45.75	3.92	16.81
	(N = 12)	*SD*	0.97	24.18	1.83	8.10
	Experimental	*M*	2.25	95.25	16.00	37.83
	(N = 12)	*SD*	0.87	15.73	4.13	6.40
Total	Control	*M*	0.61	25.96	3.84	10.13
	(N = 96)	*SD*	0.69	14.76	1.96	5.32
	Experimental	*M*	2.58	66.85	10.04	26.45
	(N = 96)	*SD*	0.97	22.68	3.30	8.19

Table 3. Cohen's *d* Effect Sizes per Variable Across Three Studies

Study	Teachers	Task Selection	Task Representation	Task Adaptation	Total PCK	Student Correct Trials
Ward et al. (2015)	**OH1**	3.58	4.09	2.19	4.85	3.01
	OH2	3.46	1.93	2.61	2.17	1.78
	NC1	1.63	2.43	3.78	2.88	3.67
	NC2	1.17	2.40	2.91	2.87	2.94
Sinelnikov et al. (2015)	**AL1**	2.59	1.06	1.04	1.11	1.85
	AL2	0.93	1.37	1.53	1.46	1.67
Iserbyt et al. (2015)	**BL1**	2.89	1.70	0.74	1.77	1.53
Mean ES		**2.32**	**2.48**	**2.11**	**2.44**	**2.35**