Check for updates

A systematic review and meta-analysis of intervention studies on mathematics and science pedagogical content knowledge

Tatsushi Fukaya*, Daiki Nakamura, Yoshie Kitayama and Takumi Nakagoshi

Graduate School of Humanities and Social Sciences, Hiroshima University, Higashihiroshima, Japan

Pedagogical content knowledge (PCK) has been considered as professional knowledge that teachers need to effectively instruct students. Empirical studies have been conducted to enrich teachers' PCK through interventions such as professional development programs. In this study, we conducted a systematic review and meta-analysis of studies about PCK interventions and their quantitative evaluation among mathematics or science teachers at the elementary and secondary education levels. We identified 101 effect sizes (42 for multiple group comparison designs and 59 for multiple time point comparison designs) and found that the interventions had a positive effect in both designs. The results of the meta-regression analysis showed that in the multiple group design, the effects differed by intervention target, with higher scores for the PCK intervention than for the CK-only intervention, and in the multiple time point design, the effects differed by subjects treated, with higher scores for the science materials than for the mathematics materials. These results not only demonstrate the average effectiveness of previous studies aimed at improving teachers' PCK but also provide insights into effective designs for professional development methods that promote PCK acquisition.

Systematic review registration: https://osf.io/vf4hq/?view_only=068483e4e82c42cd994e2c8174bd0a64.

KEYWORDS

pedagogical content knowledge, intervention, meta-analysis, systematic review, meta-regression analysis

Background of the study

The influence from teachers in the classroom may be one of the most important factors affecting student learning (Darling-Hammond, 2000; Hattie, 2009; Wayne and Youngs, 2003). A review of rigorously conducted studies examining the impact of teacher training programs on student achievement reported that enhancing teachers' instructional expertise increased student achievement by the 21st percentile (Yoon et al., 2007). Research on teacher quality and its connection to student learning (also known as teaching effectiveness) has been conducted from the 1970s (Brophy and Good, 1986; Doyle, 1977; Rosenshine, 1976) to the 2000s (Darling-Hammond, 2000; Seidel and Shavelson, 2007; Wayne and Youngs, 2003). Among these studies, since the flourishing of cognitive psychology in the 1980s, research has been particularly focused on teacher knowledge (Ball et al., 2001; Fennema and Franke, 1992; Hogan et al., 2003).

Currently, teacher professional knowledge is considered to be one of the core competencies of teachers (Baumert and Kunter, 2013). Within this knowledge, pedagogical content knowledge (PCK) proposed by Shulman (1986, 1987), which conceptualizes teacher-specific knowledge, has received particular attention. PCK is defined as follows: "It represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction" (Shulman, 1987, p. 8). Shulman (1986) posited that PCK comprises two elements. The first is knowledge of instructional representations, which is considered to be the knowledge that effectively explains certain subject matter content. The second is knowledge of learners, which is the knowledge to judge what misconceptions or bugs students show about a subject's content and how difficult (or easy) the content is.

Since Shulman's original proposition, there have been efforts to expand upon the components of PCK, as evidenced by reviews conducted by Chan and Hume (2019) for science and Depaepe et al. (2013) for mathematics. Notably, mathematical knowledge for teaching (MKT) is a well-known example of such a concept. MKT pertains to the mathematical knowledge that educators require to effectively teach mathematics and is primarily composed of content knowledge (CK) and PCK (Ball et al., 2001). One of the distinguishing features of MKT is its subdivision of the CK essential to teachers into three subcategories: common content knowledge, specialized content knowledge, and horizon content knowledge. Additionally, aside from the two elements proposed by Shulman, knowledge of content and curriculum are also regarded as elements of PCK within the context of MKT. The concept of curriculum knowledge has also been proposed as a PCK element in other studies (e.g., Kind, 2009).

Great strides have been made, not only conceptually but also methodologically, toward attempting to examine teachers' PCK empirically since Shulman's proposal. Although qualitative studies based on limited classroom observations have been conducted (e.g., Alonzo and Kim, 2016; Oh and Kim, 2013; Seymour and Lehrer, 2006), quantitative studies have also been carried out with numerous participants through test sets measuring PCK in various domains of a subject since the 2000s. Ball and colleagues have employed MKT as a framework for empirical validation, and multiple quantitative studies have been reported to date (Hill et al., 2004, 2005, 2008). Additionally, tests designed for middle school mathematics (Baumert et al., 2010), biology, and physics (Großschedl et al., 2015; Sorge et al., 2019) have been developed. Large-scale projects such as the Cognitive Activation in the Classroom (COACTIV) project, conducted in Germany as a complementary survey to the PISA program (Baumert et al., 2013), and the Teacher Education and Development Study in Mathematics (TEDS-M) project, an international survey organized by the International Association for the Evaluation of Educational Achievement (IEA; Blömeke and Kaiser, 2014), have also been undertaken.

Quantitative research utilizing the PCK test has demonstrated that students achieve higher academic results when taught by teachers who exhibit greater PCK scores (Callingham et al., 2016; Keller et al., 2016; Lenhart, 2010). A longitudinal study was conducted in Germany as part of PISA 2003, with 4,353 9th grade

students and 181 mathematics teachers participating (Baumert et al., 2010). The analysis revealed a positive relationship between teachers' PCK and students' mathematics performance at grade 10, with PCK having a larger impact on students' performance than CK [it should be noted, however, that studies such as those by Fauth et al. (2019) and Förtsch et al. (2018) reported no significant correlation between teachers' PCK and students' academic performance].

While PCK research has made progress, there have also been criticisms of the concept (see Depaepe et al., 2013 for a summary). For example, some researchers criticize that PCK is too static a view of teacher professionalism, and view PCK as more dynamic and more closely tied to actual classroom behaviors (Bednarz and Proulx, 2009; Mason, 2008; Petrou and Goulding, 2011). Cochran et al. (1993) proposed the concept of pedagogical content knowing (PCKg) instead of PCK and emphasized its dynamic nature. Although the authors acknowledge that the criticisms of PCK are important perspectives, we believe that they do not immediately invalidate the findings of PCK research. In response to the criticism of PCK static nature, recent studies have examined the relationship between static PCK and more dynamic teaching acts. These studies have found that PCK and teaching behaviors are positively related; namely, having rich PCK enables teachers to assess students' subject-based thinking and to provide effective instructions (Dreher and Kuntze, 2015; Kulgemeyer et al., 2020; Meschede et al., 2017). Thus, we suggest that PCK, even when measured statically, is important knowledge that serves as a basis for the dynamic performance of teachers.

Interventions for PCK and the purpose of this study

Considering the significant role of PCK, there have been numerous efforts to enhance the acquisition of PCK among teachers. Evens et al. (2015) conducted a systematic review of both quantitative and qualitative PCK intervention studies and found that most of these studies reported positive outcomes, indicating that the interventions had a beneficial effect on PCK development. The authors also identified five sources of PCK, namely:

(1) Educational experience: Having experience teaching actual students. In particular, early career teaching experiences are believed to promote the development of PCK (e.g., Simmons et al., 1999).
(2) PCK courses: Interventions that directly target PCK, such as thinking about what kind of misconceptions students might have and how teachers can address them (e.g., Haston and Leon-Guerrero, 2008).
(3) CK: Interventions targeting CK have also been used because having accurate and rich CK on the subject is a prerequisite for having PCK (e.g., Brownlee et al., 2001).
(4) Collaboration with peers: Collaboration among teachers, such as working together with other teachers to develop lessons and giving advice to each other (e.g., Dalgarno and Colgan, 2007).
(5) Reflection: Reflecting on one's own experiences and gaining new insights (e.g., Kenney et al., 2013).

Of these, educational experiences, PCK courses, collaboration with colleagues, and reflection were the methods most frequently mentioned in the studies, while CK was the least frequently mentioned (Evens et al., 2015).

Previous reviews have suggested the effectiveness of PCK interventions based on most studies reporting positive outcomes (e.g., Depaepe et al., 2013; Evens et al., 2015; Kind, 2009). However, these reviews failed to quantitatively measure the degree of effectiveness of such interventions. Given the problem that narrative reviews, in which researchers subjectively summarize the results of studies, lead to subjective interpretation of results (Borenstein et al., 2009), there is a need to quantify the effects of PCK interventions using meta-analysis, which can statistically integrate the results of prior studies and derive objective results. Although the development of quantitative studies examining the effects of interventions using the PCK test set has made it possible to conduct meta-analyses, no attempts have been made so far to integrate the results of the effects of PCK interventions.

To address the issue identified in previous studies that only conducted narrative reviews on PCK interventions, the first research question of this study was established: How effective have PCK interventions conducted in previous studies been? Specifically, we conduct a meta-analysis of intervention studies that have used the PCK test set to evaluate the effects of interventions on teachers of mathematics and science at the primary and secondary education stages. Previous studies on the PCK test have mainly focused on mathematics and science (Ball et al., 2008), and this study targets the same subject areas in primary and secondary education. We will include studies that use pre- and post-test comparisons at multiple time points, and multiple group comparisons such as experimental and control groups. As these are different study designs, we will integrate effect sizes based on these designs separately.

Furthermore, among studies that have conducted PCK interventions, some have reported positive effects (Rosenkränzer et al., 2017; Suma et al., 2019), while others have disclosed no effects (Luft and Zhang, 2014; Smit et al., 2017). This suggests that there may be variability in the effectiveness of PCK interventions across different studies. Therefore, this study examines the second research question: If high heterogeneity in effect sizes is observed, what factors explain this heterogeneity? Investigating this question will likely provide insights into the characteristics of effective intervention designs and materials that can promote the acquisition of PCK more effectively. To answer the second question, we also conduct meta-regression analysis examining the effects of moderators that may explain the variability of effect sizes for variables with a certain number of effect sizes. These moderators include teacher-related differences such as pre-service or in-service status and subject area and intervention-related differences such as intervention targets and duration.

Method

Recently, there has been an emphasis on the adoption of methods that improve transparency and reproducibility in conducting systematic reviews and meta-analyses, as highlighted by Polanin et al. (2020). The current study adheres to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA; Pigott and Polanin, 2020) guidelines, which prioritize transparency and reproducibility in systematic reviews.

This study was conducted as part of a project focusing on quantitative studies of PCK, which included two groups of studies: (1) a survey study examining the association between PCK and other teachers' and students' variables and (2) an intervention study evaluating the efficacy of educational interventions in enhancing PCK (this study). While the integrative statistical analyses of the survey and intervention studies were conducted separately, the literature search and coding were conducted simultaneously without distinguishing between the two.

Review protocol

Prior to conducting the literature search, a review protocol was created and registered with the Open Science Framework (OSF). The protocol is available in Appendix A at the following URL: https://osf.io/vf4hq/?view_only=068483e4e82c42cd994e2c8174bd0a64. Next, the databases and search strings for the literature search were determined. Finally, we determined the procedures for screening titles and abstracts, coding the text, and performing the meta-analysis. However, these were modified as necessary during the course of the work.

Inclusion and exclusion criteria

This study employed the following seven inclusion and exclusion criteria.

Population
The population of interest in this study was teachers responsible for mathematics and natural science education in primary and secondary schools or students pursuing relevant teaching licenses. Thus, studies that targeted teachers in charge of subjects other than mathematics and natural sciences in primary and secondary education, such as early childhood education, higher education (university), and special-needs schools, were excluded.

PCK
Included in this study are studies that quantitatively measured PCK through tests or assignments, with scores treated as a continuous variable. However, studies that did not conduct quantitative scoring, even if they included PCK tasks, as well as qualitative studies consisting solely of episodes and qualitative analysis were excluded. Furthermore, the respondents in the studies were pre- or in-service teachers, and responses measured through third-party evaluations were not included.

PCK is a unique category of teacher knowledge in that it is a mixture of CK and pedagogical knowledge (PK). Therefore, studies that only measured CK or PK, despite being referred to as PCK, were excluded. Additionally, studies that employed abstract Likert-type questionnaires that lacked specific subject content were also excluded. However, studies using MKT, the test set developed by Ball and Hill for elementary school teachers (Hill et al., 2004, 2005,

2008), were included because it consisted of both CK and PCK items. Nonetheless, studies that combined CK and PCK scores into one score were excluded from the analysis. It should be noted that what is being discussed here is the criterion for the measurement of PCK and not the intervention. In other words, if PCK was accurately measured, studies that only intervened in CK were also included herein.

Study design

For this study, we included intervention studies that compared PCK scores using multiple measurements, such as studies with an experimental group that received an intervention and a control group that did not, or studies that measured PCK scores at pre-intervention and post-intervention time points. We excluded review articles or studies that relied on secondary data, as well as studies that measured PCK but only described its characteristics without conducting intervention. Studies explicitly aimed at enhancing professional knowledge other than PCK were excluded even if they conducted interventions (Harr et al., 2014, 2015). Studies measuring developmental changes in PCK across university years were included, as they can be considered broad intervention studies reflecting the effects of university education (e.g., Blömeke et al., 2014).

Statistics

The studies that provided information on the mean and standard deviation (or standard error and sample size) for each group or pre- and post-period were included in our analysis. Studies that did not report the necessary statistics were not automatically excluded, but we attempted to contact the authors to obtain the statistics or raw data. When the contact information for the first or corresponding author was not available online or if the author did not respond or declined to provide information, the study was excluded.

Time frame

The review included studies that were published after the proposal of PCK by Shulman in 1986.

Publication status

To avoid publication bias, the review included non-published papers such as conference proceedings and doctoral dissertations in addition to papers published in academic journals. However, conference presentations, master's theses, and books were excluded from the analysis. Books were excluded from the analysis due to their lack of original data and potential overlap with journal papers.

Language

The study included papers written in English and Japanese, the latter being the native language of the research group.

Literature search, screening, and data extraction

A systematic review was conducted to identify intervention studies examining the effectiveness of an education program to promote PCK acquisition among pre- or in-service teachers, in addition to survey studies. To this end, the literature search initially collected both intervention and survey studies on mathematics and science, which were later classified separately after screening.

Search targets

The primary search for relevant studies was conducted using online databases. To supplement this search, manual searches were conducted on the reference lists of PCK review articles and books with PCK in the title to ensure that all relevant references were identified.

Databases used

To gather a comprehensive collection of studies, we utilized representative databases (Web of Science) and a search engine (Google Scholar), as well as ProQuest Dissertations and Theses, which enables the exploration of doctoral dissertations, and Citation Information by the National Institute of Informatics (CiNii) for Japanese-language articles.

Search string

Electronic searches in Web of Science, ProQuest, and Google Scholar were conducted using the following character strings: ("pedagogical content knowledge" OR PCK OR "mathematical knowledge for teaching" OR "content knowledge for teaching mathematics") AND (science OR sciences OR scientific OR chemistry OR chemical OR biology OR biological OR physics OR mathematics OR mathematic OR mathematical OR math) NOT ("technological pedagogical content knowledge" OR TPCK OR "technological pedagogical and content knowledge" OR TPACK OR "pedagogical technology integration content knowledge" OR PTICK). In CiNii, the following strings were used: ("pedagogical content knowledge" OR PCK OR "mathematical knowledge for teaching" OR "content knowledge for teaching mathematics" OR "授業を想定した教材内容知識" OR "授業を想定した教科内容の知識").

Procedures

Figure 1 depicts the search process and its outcomes, which was conducted according to the PRISMA 2020 guidelines for systematic reviews (Page et al., 2021). The searches were conducted between April 13 and 23, 2021, using the Web of Science, Google Scholar, ProQuest, and CiNii databases. The first and fourth authors searched Google Scholar and CiNii, while the second and third authors searched Web of Science and ProQuest. The search results were cross-checked by two authors. Additionally, we conducted a manual search of the reference lists of relevant review articles and books. The search terms used in the database were also used to scrutinize the reference lists, leading to the identification of 33 papers. The searches

FIGURE 1
Search and results flow chart.
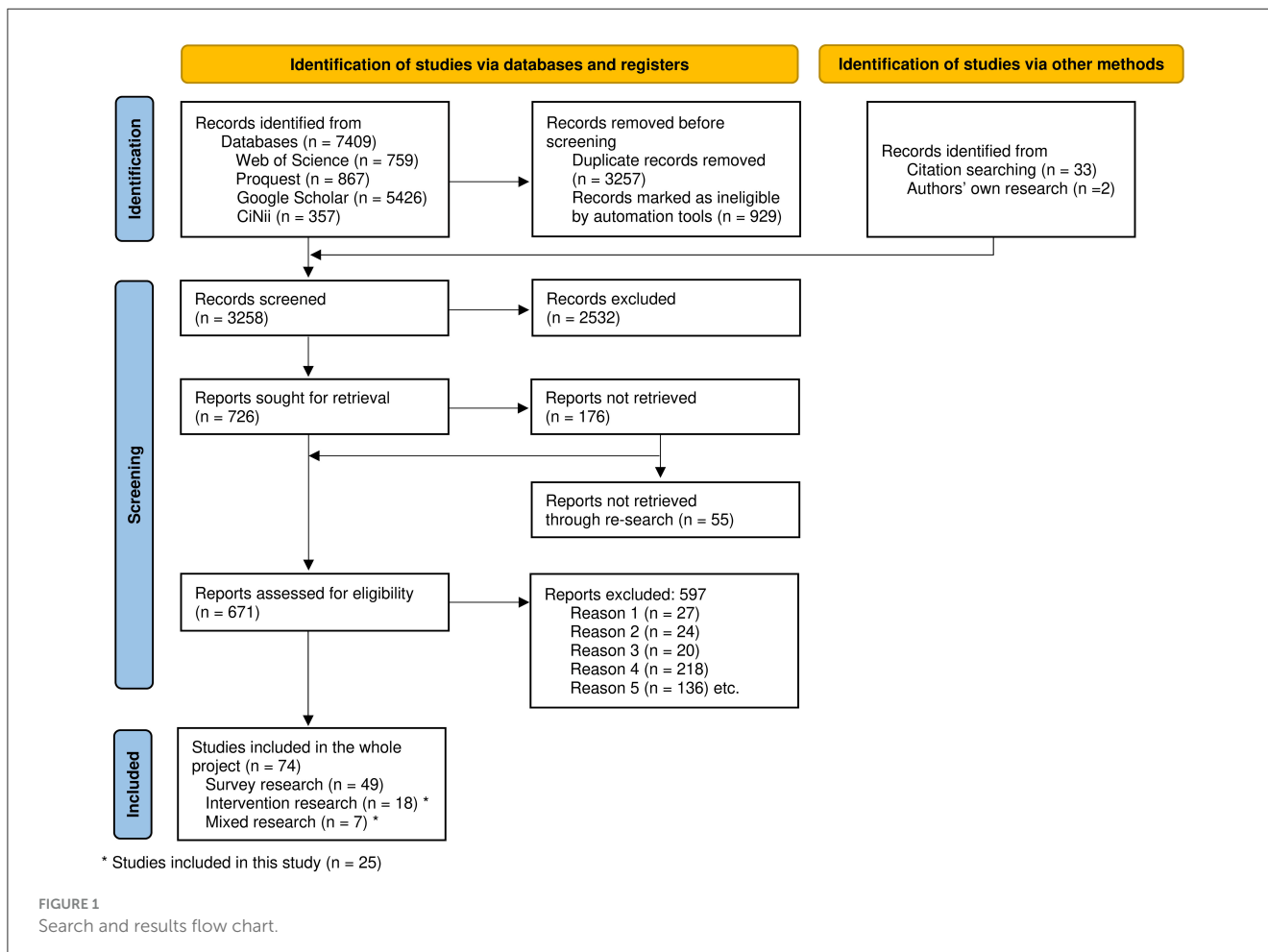
resulted in a total of 7,409 hits; 3,257 duplicates and 929 ineligible records were removed, and 3,258 references were screened by adding the 33 references found through the hand search.

From May 21 to October 6, 2021, the titles and abstracts were reviewed by the members of the research team. After 10 papers were screened as practice, the remaining papers were distributed among the four authors for screening. Any uncertainties or disagreements were discussed by all team members, and the final decision was made by consensus. A total of 726 references were chosen for full-text preservation, of which 176 references were excluded because they were either inaccessible or behind a paywall in databases that were not covered by the authors' university contract. However, considering that the authors' university entered into a new database contract in June 2023, we reattempted to access the full texts of these 176 references from October 1 to 11, 2023. As a result, we were able to access the full texts of 121 of the 176 references, reducing the number of reports not retrieved to 55; thus, we conducted a full-text screening of 671 references.

From October 6, 2021 to January 14, 2022, we screened the full texts. After practicing the screening procedures for 15 references together, the four authors divided the references among themselves and worked on them separately. Then, all members discussed the cases in which there were doubts, and the final value was determined by consensus. The 121 references that were added later

were screened by the first author between October 11 and October 30, 2023. Consequently, 598 articles were deemed ineligible for the following reasons: not in English or Japanese (Reason 1, $n = 27$), not empirical papers (Reason 2, $n = 24$), not the target sample (Reason 3, $n = 20$), PCK not quantitatively measured (Reason 4, $n = 218$), PCK not scored as a continuous variable (Reason 5, $n = 136$), and others. The remaining 74 articles were subject to coding for main texts.

From January 14 to February 4, 2022, the first round of coding was carried out. As we mentioned, in this study, we conducted a systematic review of both intervention studies that investigated the impact of interventions targeting PCK and survey studies that examined the relationship between PCK and other teachers' and students' related variables. Therefore, we first coded the type of studies to distinguish between intervention and survey studies. Between February 4 and April 8, 2022, the main coding process was carried out. When the coding method was confirmed with all members, three papers were coded as a practice to verify the coding method. Then, the papers were divided between the first and fourth author and the second and third authors, who discussed and coded the papers in pairs. Finally, the first, second, and third authors met to discuss and confirm the results of the coding. The literature added later was coded by the first author between February 14 and 29, 2024, during which nine articles met the screening criteria and five studies were coded. The data that were coded by the second and third

authors were independently coded by the first author to determine the inter-rater agreement. Upon coding 11 studies and 42 effect sizes, the range of agreement was 88% to 100%, with kappa coefficients ranging from 0.81 to 1.00 ($M = 0.98$, $SD = 0.05$). Discrepancies were resolved through discussion between the first and second authors.

Six papers lacked the necessary statistics for the meta-analysis. The authors attempted to contact the authors of all six papers for which they could confirm the first author's contact information on the internet and received responses from three of them, one of whom provided raw data. This study (Luft and Zhang, 2014) was included in the meta-analysis. Note that when a longitudinal study design with three or more time points was used (e.g., Luft and Zhang, 2014), multiple effect sizes were extracted based on the difference between Time 3 and Time 2 in addition to the difference between Time 2 and Time 1. For studies with three or more groups (e.g., Rosenkränzer et al., 2017), we assigned experimental and control groups based on the description in the paper. As a result, of the 74 studies, 49 were survey studies; therefore, 25 intervention studies were included in this study.

Coding variables

The main coding variables of the study were as follows.

Variables related to the surface characteristics of the study: (1) the year of publication of the article, (2) the country in which the study was conducted, and (3) the publication status of the study.

Characteristics related to the sample: (1) teacher status (pre- or in-service teachers), (2) school type of the teachers (classified according to the notation in the paper), (3) subject specialization of the teachers, and (4) years of teaching experience of the teachers.

Characteristics related to the measurement of PCK: (1) elements of PCK (knowledge of instructional representations, knowledge of learners, or others), (2) subject domain (number and computation, functions, physics, chemistry, etc.), (3) test format (multiple choice, free description, or others), (4) level of representations measured (memory level or transfer level), and (5) number of PCK test items. Note, however, that for a majority of the effect sizes (92%), information regarding the level of representation (transfer level vs. memory level) assessed in the PCK test was unavailable and hence was not considered in the subsequent analyses.

Characteristics related to the intervention of PCK: (1) intervention target (CK, knowledge of instructional representations, knowledge of learners, or others) and (2) intervention method. For the intervention methods, based on a review by Evens et al. (2015), we addressed four activities as follows. The first is collaborative activities that involve peer collaboration, such as discussions among participants. The second is practical activities that involve one's own or others' educational practices, such as mock classes or classroom observations. The third is theoretical activities that provide theoretical knowledge about subjects or education, such as lectures by researchers. Fourth is reflection, which involves ex post discussions of one's own or others' practice. (3) Intervention duration. The duration of the intervention was coded as follows based on the description in the article: several hours (1–6 h), several days (6–10 h), several weeks (10–30 h), several months (30 h or more per month), and years (a period of more than 1 year). (4)

Study design (one group pre-post comparison design, or two groups post comparison design, and others).

Statistics: (1) mean, SD, sample size, and others for the posttest of the experimental group, (2) mean, SD, sample size, and others for the posttest of the control group, (3) mean, SD, sample size, and others for the pretest of the experimental group.

Analytical procedures

We employed the dmerat (version 0.0.9000), meta (version 6.0-0), and metafor (version 3.8-1) packages for R (version 4.2.1) to conduct statistical analyses. The R script containing all the reported analyses (Appendix B) and the raw data file used in the analyses (Appendix C) can be accessed from the aforementioned URL of OSF.

Criteria for integration of effect sizes and the model to be employed

To calculate the effect size, Hedges' g was used, which represents the effect size between two independent groups. Hedges' g is calculated by adding a small sample correction to Cohen's d. The effect of the correction is particularly large when the sample size is small ($n < 20$). Note that in studies with multiple time point comparisons, the data between time points are not independent because the same individual was measured several times. To calculate the effect size using paired data, we need not only the mean and standard deviation of each time point but also the correlation coefficient between time points, which was not reported in most studies. Therefore, in this study, Hedges' g was used as the effect size even for studies with designs based on multiple time point comparisons.

The effect sizes obtained were synthesized using the random-effects model to estimate the overall effect size of the study. Borenstein et al. (2009) state that it is generally appropriate to apply a random-effects model, unless the researchers, populations, and procedures are all the same. Additionally, Harrer et al. (2021) point out that in the field of social sciences, a random-effects model is often employed because heterogeneity in studies is always assumed; therefore, it is customary to always use a random-effects model.

Methods of analysis in the integration of effect sizes

The study used four methods to calculate the integrated effect size: (1) baseline method (simple random-effects model), (2) trim-and-fill method, (3) outlier removal method, and (4) three-level model method. The first method (baseline method) integrated all effect sizes based on a random-effects model (Borenstein et al., 2009). The second method (trim-and-fill method) was employed when the asymmetry test (Egger test) was significant (Egger et al., 1997), where pseudo data were added to create a symmetrical distribution of the effect size around the mean, followed by integration. The outlier removal method excluded effect sizes outside the 95% confidence interval of the mean effect size of the baseline as outliers before integration (Viechtbauer and Cheung, 2010).

The second and third methods address the issues of asymmetry or outliers in the collected effect sizes. However, a related issue caused by the nested structure of effect sizes has been recognized by researchers in the field of meta-analysis. In many meta-analyses, effect sizes are nested within studies, resulting in non-independent data when multiple effect sizes are reported within a study. To address this issue, the fourth method (the three-level model method) was employed to calculate integrated effect sizes while accounting for the nested structure of effect sizes (Harrer et al., 2021). Ignoring the nested structure of data within study clusters is known to introduce bias in effect size estimation (Becker, 2000). Therefore, the fourth method was used to integrate effect sizes while taking the nested structure into account.

Evaluating heterogeneity

To assess the heterogeneity in effect sizes across studies, we utilized the I^2 statistic (Higgins et al., 2003). The study by Higgins et al. (2003) proposed that I^2 values equal to 25%, 50%, and 75% indicate low, moderate, and high heterogeneity, respectively. We estimated the heterogeneity variance τ^2 using REML estimators (Viechtbauer, 2005). To calculate the confidence intervals for the pooled effects, we employed the Knapp-Hartung adjustment (Knapp and Hartung, 2003).

Meta-regression analysis

When high heterogeneity was observed and there were more than 10 effect sizes available for each level of the variables, the researchers conducted a meta-regression analysis to investigate the study characteristics (moderators) that caused the heterogeneity (Borenstein et al., 2009). The model for the meta-regression analysis is represented by Equation 1.

$$\hat{\theta}_k = \theta + \beta_1 x_{1k} + \ldots + \beta_n x_{nk} + \epsilon_k + \zeta_k \tag{1}$$

Equation 1 corresponds to a random-effects model in which the estimated effect size $\hat{\theta}_k$ of study k is expressed by a linear combination of the intercept θ, the product of the moderation variable x_{nk} and regression coefficient β_n, and the sampling error ζ_k. The researchers used the maximum likelihood method to estimate the models. If the regression coefficient test in the adopted model is significant at the 5% level, the moderation variable is considered to have a significant effect. In the meta-regression analysis, data containing missing values were subject to list wise deletion. Therefore, note that the number of sample size in meta-regression analysis is smaller compared to other analyses.

Results

Characteristics of the collected studies

Table 1 provides an overview of the key characteristics of the 25 studies that were included in this study. The majority of the studies were published after 2010, and there has been a noticeable increase in the number of quantitative PCK intervention studies in recent years. The United States was the most common country of publication, contributing nearly half of the total studies, with Germany being the second most common location, with five studies. The studies were predominantly published in academic journals. Approximately half of the studies involved pre-service teachers, while the other half included in-service teachers. There was an almost equal number of studies focused on mathematics and on science. The most common intervention design was a one-group pre-post comparison design, which was used in 15 studies. In contrast, a total of 10 studies used multiple groups comparison designs.

In the subsequent sections, we divided the analysis into effect sizes based on multiple group comparison designs and those based on multiple time point comparison designs and analyzed each effect size separately. Initially, we provided a summary of the descriptive features of the tests and interventions and then reported results regarding the integration of effect sizes and meta-regression analysis.

Analysis of effect sizes based on multiple group comparison design

First, we focused on the characteristics of PCK tests and interventions in the effect sizes presented in multiple group comparison designs ($k = 42$). In Table 2, we outlined the features of the PCK tests, with many studies measuring a single component of PCK ($k = 27$) and about 30% studies measuring multiple components ($k = 14$). The majority of the effect sizes came from science tests, with fewer from mathematics. This disparity was because the majority of papers that reported a significant number of effect sizes were from science studies. While half of the test formats were descriptive tests, only a small number were multiple-choice or interview tests. The number of items on the PCK test varied significantly, ranging from 2 to 58, with 5 being the most common.

Table 3 presents the characteristics of interventions used in the effect sizes. While few educational methods used in the interventions were classified in terms of specific characteristics, many of them combined multiple characteristics. Thus, in addition to the number of effect sizes classified only as specific characteristics, we reported the number of effect sizes with individual characteristics among those that combined multiple characteristics. As a result, the ratios in the intervention targets and educational methods in Table 3 exceeded 100%. Of the 42 effect sizes, 15 targeted only CK in their interventions. On the other hand, many of the studies that had multiple types of knowledge as their intervention target, directly targeted PCK as well as CK. Of these, knowledge of learners were the most frequently targeted components of PCK. The majority of interventions were designed using multiple methods, with lectures by researchers being the most common (29 cases), followed by collaborative activities (27 cases), practical activities (15 cases), and reflection (12 cases). There were no studies that conducted interventions over short periods such as hours or days; the most common duration for interventions was in weeks. More than 20% of each variable was unknown, and many of the studies did not specify the details of the intervention.

Furthermore, an analysis was conducted to synthesize the effect sizes obtained from studies that utilized multiple group comparison

TABLE 1 Characteristics of the sampled studies (n = 25).

Variable	Number of cases	Ratio
Year of publication		
1980s	0	0.0%
1990s	0	0.0%
2000s	5	20.0%
2010s and after	20	80.0%
Country/Region		
USA	12	48.0%
Germany	5	20.0%
Asia (Singapore, Indonesia)	4	16.0%
Oceania (Australia)	1	4.0%
Africa (South Africa)	1	4.0%
Combined (Germany, Switzerland, etc.)	2	8.0%
Publication status		
Academic journals	17	68.0%
Proceedings	3	12.0%
Dissertation	5	20.0%
Teaching experience of the sample		
Pre-service teachers	13	52.0%
In-service teachers	12	48.0%
School types of the sample		
Elementary/primary school	7	28.0%
Junior-high school	2	8.0%
Senior-high school	3	12.0%
Middle school	5	20.0%
Combined	4	16.0%
Unknown	4	16.0%
Studied subjects		
Mathematics	10	40.0%
Biology	3	12.0%
Chemistry	3	12.0%
Physics	2	8.0%
Science (combined)	2	8.0%
Unknown	5	20.0%
Research design		
One-group pre-post comparison design	15	60.0%
Two-groups post comparison design	6	24.0%
Others (More than two groups and two time points comparison design)	4	16.0%

TABLE 2 Characteristics of the PCK test in effect sizes reported in multiple group comparison designs (k = 42).

PCK test characteristic	Number of cases/values	Ratio
Elements		
1. Knowledge of instructional representations	9	21.4%
2. Knowledge of learners	12	28.6%
3. Other Elements	6	14.3%
4. Combined (1, 2)	2	4.8%
5. Combined (1, 2, 3)	11	26.2%
6. Combined (other than 1, 2)	1	2.4%
7. Unknown	1	2.4%
Subject		
1. Mathematics	9	21.4%
2. Natural sciences (biology, chemistry, physics)	25	59.5%
3. Unknown	8	19.0%
Test format		
1. Multiple choice test	5	11.9%
2. Descriptive test	20	47.6%
3. Other formats	7	16.7%
4. Combined	5	11.9%
5. Unknown	5	11.9%
Number of items		
Min	2	
Max	58	
M	7.47	
SD	10.67	
Median	5	

size and its 95% confidence interval, respectively. The diamond at the bottom of the plot indicates the overall effect size and its 95% confidence interval, as aggregated by the random-effects model. The vertical line at the center of the plot represents an effect size of 0, which serves as the reference line for determining the significance of the effect sizes. In Figure 2, the left edge of the diamond crosses this reference line, indicating that the effect of the intervention is significant.

Table 4 presents the results of effect size integration for each of the four models. The table sequentially displays the sample size (*k*), the effect size (*g*) with its 95% confidence interval, and the significance probability (*p*). The 95% prediction interval (PI) indicates the range within which future observations are expected to fall. This interval predicts the range of effect sizes that may be observed in subsequent studies based on the results obtained from the meta-analysis (Borenstein et al., 2009). The I^2 statistic measures the extent of heterogeneity in the meta-analysis. As shown in Table 4, the integration of effect sizes using the random-effect

designs. The forest plot is shown in Figure 2, and the statistics based on the four models are shown in Table 4. In the forest plot, the effect sizes extracted from the included studies are displayed in ascending order. The squares and the horizontal lines represent each effect

TABLE 3 Characteristics of the PCK intervention in effect sizes reported in multiple group comparison designs (k = 42).

PCK intervention characteristic	Number of cases/values	Ratio
Targets		
1. CK only	15	35.7%
2. Others only	1	2.4%
3. Combined (including CK)	17	40.5%
4. Combined (including knowledge of learners)	19	45.2%
5. Combined (including knowledge of instructional representations)	15	35.7%
6. Combined (including others)	5	11.9%
7. Unknown	7	16.7%
Educational methods		
1. Theoretical activities only	1	2.4%
2. Practical activities only	1	2.4%
3. Combined (including collaborative activities)	27	64.3%
4. Combined (including practical activities)	15	35.7%
5. Combined (including theoretical activities)	29	69.0%
6. Combined (including reflection)	12	28.6%
7. Unknown	11	26.2%
Durations		
1. Hour	0	0.0%
2. Day	0	0.0%
3. Week	18	42.9%
4. Month	11	26.2%
5. Year	4	9.5%
6. Unknown	9	21.4%

model indicated that the intervention in the experimental group led to higher performance on the PCK test than the control group, with an expected effect size of $g = 0.74$ (95% CI [0.51, 0.97]). However, high heterogeneity was confirmed between the studies, with an I^2 value of 84% (95% CI [79%, 88%]).

The Egger test yielded a significant result [$t_{(40)} = 2.12$, $p < 0.05$], indicating asymmetry in the effect size distribution. Thus, the trim-and-fill method was employed, which yielded a lower expected effect size of $g = 0.59$ (95% CI [0.33, 0.85]) than the baseline model. Additionally, an analysis that excluded outliers based on confidence intervals yielded an expected effect size of $g = 0.67$ (95% CI [0.48, 0.85]). Although the expected effect size was slightly lower than the baseline model, the difference was not substantial. Finally, the three-level model produced a mean effect size of $g = 0.63$ (95% CI [0.28, 0.98]). Although the effect size was smaller than the other two models, it may be attributed to the study-level variation caused by large effect sizes reported by some studies such as Rosenkränzer et al. (2017), as indicated by estimated variance components of τ^2 Level 3 = 0.17 and τ^2 Level 2 = 0.23. This indicates that I^2 Level 3

= 36% and I^2 Level 2 = 48% of the variation were attributed to intercluster and intracluster heterogeneity, respectively. The three-level model provided a better fit than the two-level model ($\chi^2 =$ 11.93, $p < 0.001$).

To account for the substantial heterogeneity in the effect sizes observed across studies, a meta-regression analysis was conducted to identify the variables that may explain these differences. The study included independent variables of teachers' countries (USA teachers vs. German teachers), teachers' teaching experience (in- vs. pre-service teachers), PCK test components (single vs. multiple components), intervention targets (CK only vs. multiple elements including PCK), and intervention duration (less than weeks vs. longer than months) for the more than 10 samples under each condition. Table 5 presents the results of the meta-regression analyses of the variables accordingly. The effect of intervention targets (CK plus PCK Intervention) was significant (Table 5), indicating that interventions targeting PCK were found to enhance the effectiveness of the intervention compared to interventions that focused on CK alone ($p = 0.01$). On the other hand, differences in teachers' countries, teaching experience, PCK test components, and intervention duration did not have a significant effect on the results.

Analysis of effect sizes based on multiple time point comparison design

We also summarized the characteristics of PCK tests and interventions in the effect sizes reported in multiple time point comparison designs. Table 6 shows the characteristics of the PCK tests. The data showed that approximately 60% of the PCK tests evaluated a single element (37 cases), while the rest assessed multiple elements (22 cases). Science was the main subject, but mathematics accounted for about 20% of the PCK tests, similar to those in the multiple group comparison design studies. The most common test format was descriptive tests, and there were few multiple-choice or other types of tests, such as interviews. The number of PCK items varied considerably from 2 to 58, with 5 being the most commonly used.

Table 7 provides a summary of the characteristics of interventions used in the targeted effect sizes. On the intervention targets, 11 cases solely focused on CK, while the remaining interventions targeted multiple types of knowledge, with many of them directly addressing both PCK and CK. Of these, knowledge of instructional representations and knowledge of learners were addressed in 24 cases each. Furthermore, all interventions utilized more than one educational method, with collaborative activities being the most commonly used (39 cases). The second most common were theoretical activities led by researchers (30 cases), followed by practical activities such as class observation (25 cases) and teacher reflection (16 cases). The majority of interventions lasted for weeks, but some were conducted over several months or years. As with the group comparison design, many variables were unknown, such as the intervention targets in 34% of the effect sizes.

Furthermore, an analysis to combine the effect sizes obtained from the studies in the multiple time point comparison design was conducted. The forest plot for this analysis is presented in Figure 3, and the statistics for the four models are displayed in
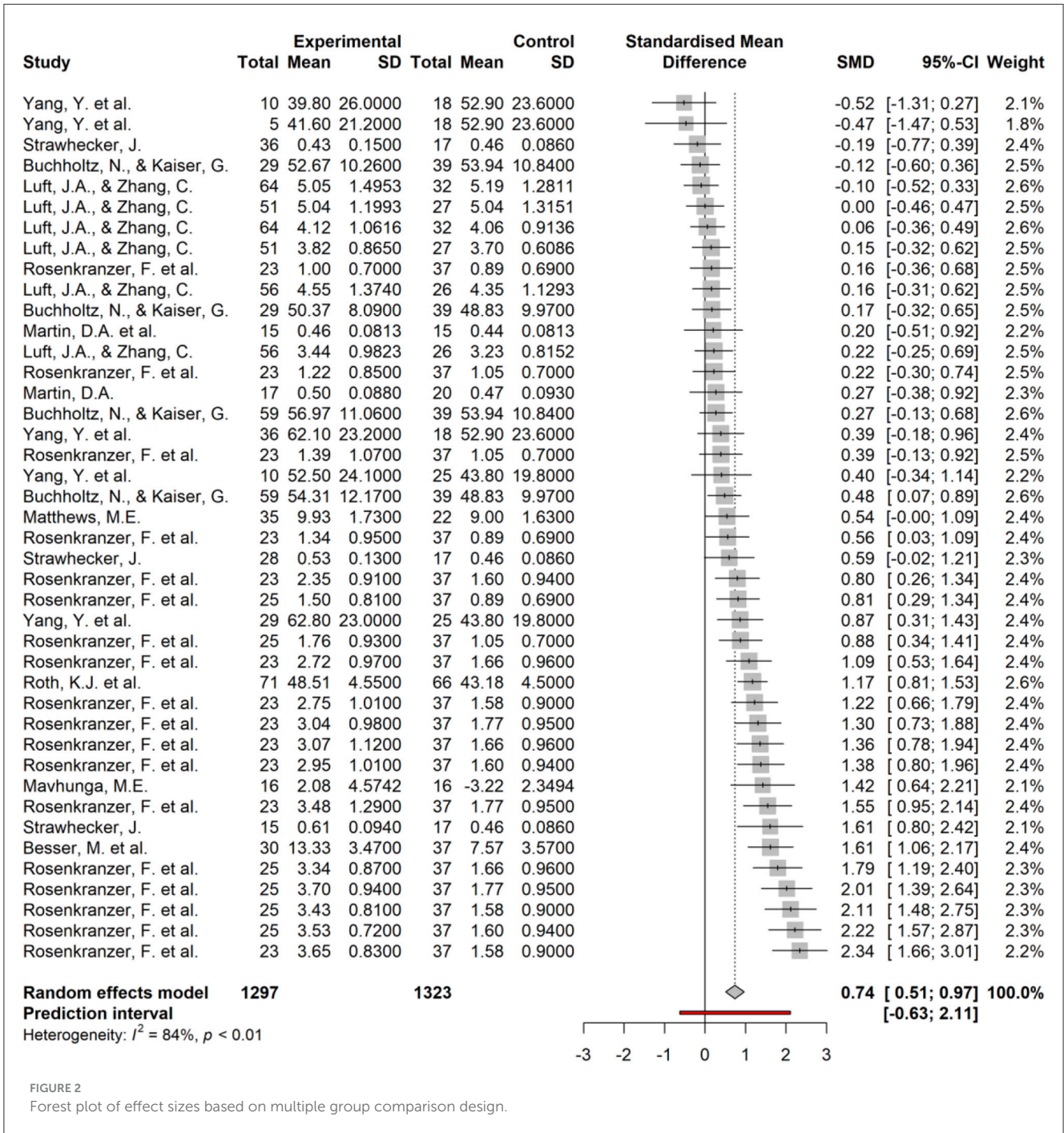
FIGURE 2
Forest plot of effect sizes based on multiple group comparison design.

TABLE 4 Results of the integration of effect sizes of multiple group comparison design based on four models.

Model	k	g [95% CI]	P	95% PI	I^2 [95% CI]
Baseline	42	0.74 [0.51; 0.97]	<0.0001	−0.63; 2.11	84% [79%, 88%]
Trim-and-fill	46	0.59 [0.33; 0.85]	<0.0001	−1.06; 2.24	87% [84%, 90%]
Outlier removal	30	0.67 [0.48; 0.85]	<0.0001	−0.15; 1.49	67% [51%, 77%]
Three-level model	42/11	0.63 [0.28; 0.98]	<0.001	−0.71; 1.97	L2 48%, L3 36%

TABLE 5 Results of the meta-regression analysis for the effect sizes of multiple group comparison design (k = 29).

	Estimate	95% CI	t	p
Intercept	0.62	[−0.80, 2.04]	0.90	0.37
USA teachers (vs. German teachers)	−0.36	[−1.73, 1.01]	−0.54	0.59
Pre-service teachers (vs. In-service teachers)	0.42	[−1.73, 0.89]	−0.66	0.52
CK plus PCK Intervention (vs. CK only)	0.69	[0.18, 1.21]	2.76	0.01
Longer intervention (multiple test elements) [1]	0.54	[1.60, 1.24]	1.60	0.12

1) In the data for this study, the variables of intervention duration (weeks vs. months) and PCK test elements (single vs. multiple) are perfectly confounded, in which, all studies conducted less than weeks used single element test, while all studies conducted longer than months used multiple elements test. As a result, these two factors were completely confounded in this study.

TABLE 6 Characteristics of the PCK test in effect sizes reported in multiple time comparison designs (k = 59).

PCK test characteristic	Number of cases/values	Ratio
Elements		
1. Knowledge of instructional representations	10	16.9%
2. Knowledge of learners	12	20.3%
3. Other Elements	15	25.4%
4. Combined (1, 2)	2	3.4%
5. Combined (1, 2, 3)	12	20.3%
6. Combined (other than 1, 2)	8	13.6%
7. Unknown	0	0.0%
Subject		
1. Mathematics	14	23.7%
2. Natural sciences (biology, chemistry, physics)	31	52.5%
3. Unknown	14	23.7%
Test format		
1. Multiple Choice Test	4	6.8%
2. Descriptive Test	25	42.4%
3. Other Formats	11	18.6%
4. Combined	13	22.0%
5. Unknown	6	10.2%
Number of items		
Min	2	
Max	58	
M	8.92	
SD	12.10	
Median	5	

TABLE 7 Characteristics of the PCK intervention in effect sizes reported in multiple time comparison designs (k = 59).

PCK intervention characteristics	Number of cases/values	Ratio
Targets		
1. CK only	11	18.6%
2. Combined (including CK)	28	47.5%
3. Combined (including knowledge of learners)	24	40.7%
4. Combined (including knowledge of instructional representations)	24	40.7%
5. Combined (including others)	7	11.9%
6. Unknown	20	33.9%
Educational methods		
1. Combined (including collaborative activities)	39	66.1%
2. Combined (including practical activities)	25	42.4%
3. Combined (including theoretical activities)	30	50.8%
4. Combined (including reflection)	16	27.1%
5. Unknown	20	33.9%
Durations		
1. Hour	2	3.4%
2. Day	4	6.8%
3. Week	19	32.2%
4. Month	15	25.4%
5. Year	8	13.6%
6. Unknown	11	18.6%

Table 8. As shown in Table 8, the results of random-effects model indicated an expected improvement in PCK scores from pre- to post-intervention, with an effect size of $g = 0.90$ (95% CI [0.71, 1.10]). However, the high I^2 value of 87% (95% CI [85%, 90%]) indicated significant heterogeneity among the studies. Additionally, the Egger test yielded a significant result [$t_{(51)} = 5.25$, $p < 0.001$],

indicating asymmetry in the effect size distribution. Thus, the trim-and-fill method was employed, which yielded a lower expected effect size of $g = 0.44$ (95% CI [0.18, 0.69]) than the baseline model. A subsequent analysis that excluded outliers based on confidence intervals yielded an expected effect size of $g = 0.91$ (95% CI [0.73, 1.09]), which was similar to the baseline model.

The effect sizes obtained from the three-level meta-analytic model had a mean value of $g = 0.79$ (95% CI [0.49, 1.09]). Although this effect size was smaller than that obtained from the other two models, this was likely due to accounting for variation between studies that report several high effect sizes. The estimated

variance components were τ^2 Level 3 = 0.33 and τ^2 Level 2 = 0.11, indicating that I^2 Level 3 = 68% of the variation was due to intercluster heterogeneity, and I^2 Level 2 = 22% was due to intracluster heterogeneity. The three-level model showed a better fit than the two-level model (χ^2 = 29.96, p < 0.001). The magnitude of the effect sizes differed among the models, with the trim-and-fill and three-level model analysis methods yielding lower effect sizes than the other models. This suggests the need for more studies, including those in which sufficient intervention effects were not obtained. However, all models indicated a positive effect size, indicating that the intervention was effective in improving teachers' PCK.

To account for the high level of heterogeneity observed across the studies, a meta-regression analysis was conducted to identify the variables that may explain the variations in the effect sizes. The meta-regression analysis included more than 10 studies for each condition, and the following independent variables were entered into the analysis: teachers' countries (USA teachers vs. German teachers), teachers' teaching experience (in- vs. pre-service teachers), PCK test components (single vs. multiple components), subject measured by PCK test (mathematics vs. science), intervention targets (CK only vs. multiple elements including PCK), and intervention duration (less than weeks vs. longer than months). Only the regression coefficient for the subject measured by PCK test (Science PCK vs. Math PCK) was statistically significant (Table 9), indicating that interventions had a greater effect when the subject was science compared to mathematics (p = 0.04). On the other hand, differences in teachers' countries, teaching experience, PCK test components, intervention targets, and intervention duration did not have a significant effect on the results.

Discussion

This study conducted a systematic review and meta-analysis of quantitative PCK studies to investigate the effects of teacher PCK interventions in mathematics and science education at elementary and secondary levels. The review involved searching representative research databases and collecting 101 effect sizes from 25 studies, including 42 effect sizes for multiple group comparison designs and 59 effect sizes for multiple time point comparison designs. While several studies have examined PCK interventions since their proposal by Shulman (1986), most of the studies included in this review were conducted after 2010, and studies that quantitatively examined the effects of PCK interventions using PCK tests were relatively recent. Therefore, the number of effect sizes collected was not extensive, which is a limitation of this study that will be further discussed.

When effect sizes were integrated, the expected effect size for the multiple group comparison design was 0.74 (baseline model), with higher PCK scores for the experimental group receiving the intervention than for the control group not receiving the intervention. The expected effect size for the multiple time point comparison design was 0.90 (baseline model), indicating higher

scores on the PCK test measured after the intervention compared to the scores measured a priori. This result confirms the claim in previous reviews (e.g., Depaepe et al., 2013; Evens et al., 2015; Kind, 2009) that have argued for the effectiveness of PCK interventions. However, these reviews did not quantify the extent to which PCK interventions were effective. The results of the present study show that the PCK intervention has a large effect in both designs. On the other hand, estimates based on the three-level model tended to be smaller than the baseline model, with g = 0.63 for the multiple group comparison design and g = 0.79 for the multiple time point comparison design. This is because a large number of effect sizes were reported in a single study. For example, Rosenkränzer et al. (2017) reported 18 effect sizes in each design, which included very high effect sizes. The three-level model takes into account the study level, which relatively weakens the effect of studies reporting more than multiple effect sizes. Therefore, the multilevel model can be interpreted as having lower effect sizes than the baseline model.

Comparing the two designs, higher effect sizes were found in the multiple time point comparison design than in the between-group comparison design. In many cases of multiple group comparison designs, teachers, even in a control group, participated in some learning programs (e.g., Luft and Zhang, 2014; Strawhecker, 2005; Yang et al., 2018). For example, in Roth et al. (2019), teachers in the control group participated in a program to deepen their content knowledge, as opposed to those in the experimental group who engaged in activities to analyze their own and others' classroom videos over a 1-year period. Control group teachers engaged in a variety of content-deepening learning activities that they then implemented in their classrooms. Thus, in the multiple group comparison design, the difference between the two groups tended to be smaller. On the other hand, in the multiple time point comparison design, the scores before receiving the intervention are used as the point of comparison, so the intervention effect is likely to be more pronounced.

On the other hand, there was large heterogeneity in each intervention, as expressed in the I^2 values. Therefore, in this study, we conducted a meta-regression analysis to test which factors could explain the variation in effect sizes. The analysis of multiple group comparison design revealed effects across intervention targets (CK-only vs. CK plus PCK intervention), with higher effects for the intervention including PCK as well compared to the CK-only intervention. Many studies on PCK have targeted CK based on the assumption that CK is necessary for teachers to acquire PCK (Abell, 2008; Gess-Newsome, 2013). However, studies that rigorously tested this hypothesis through experiments reported that teaching CK to teachers had minimal effects on PCK and that direct instruction in PCK was more effective (Evens et al., 2018; Tröbst et al., 2018). The results of this study can be seen as corroborating these experimental findings through meta-analysis.

On the other hand, the result with data of multiple time point comparison design showed the effect of subject measured by PCK tests, indicating that interventions had a greater effect when the subject was science compared to mathematics. This result can be attributed to educational activities specific to the science domain that are not typically conducted in

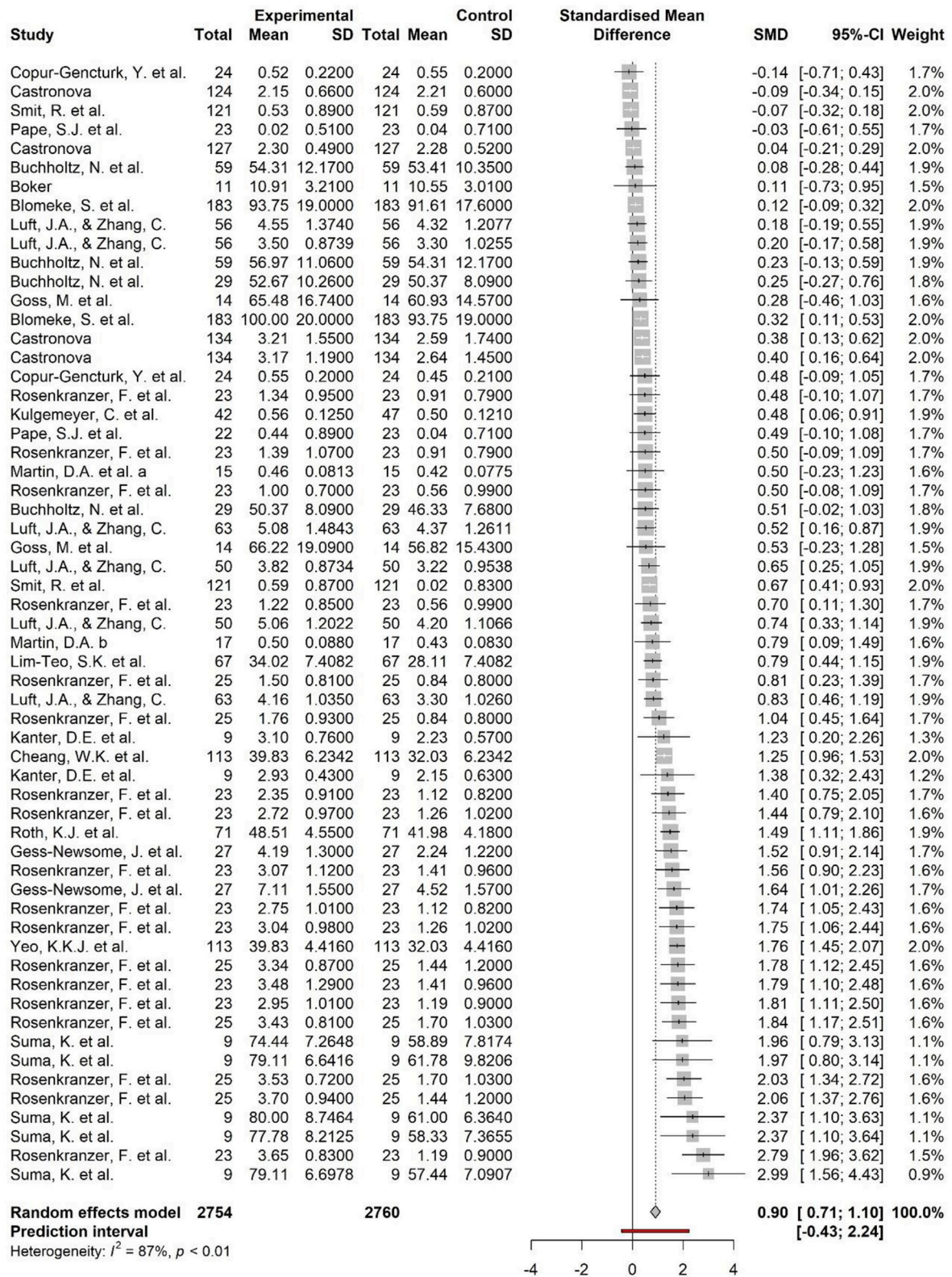| Study	Total	Experimental Mean	SD	Total	Control Mean	SD	Standardised Mean Difference	SMD	95%-CI	Weight
Copur-Gencturk, Y. et al.	24	0.52	0.2200	24	0.55	0.2000		-0.14	[-0.71; 0.43]	1.7%
Castronova	124	2.15	0.6600	124	2.21	0.6000		-0.09	[-0.34; 0.15]	2.0%
Smit, R. et al.	121	0.53	0.8900	121	0.59	0.8700		-0.07	[-0.32; 0.18]	2.0%
Pape, S.J. et al.	23	0.02	0.5100	23	0.04	0.7100		-0.03	[-0.61; 0.55]	1.7%
Castronova	127	2.30	0.4900	127	2.28	0.5200		0.04	[-0.21; 0.29]	2.0%
Buchholtz, N. et al.	59	54.31	12.1700	59	53.41	10.3500		0.08	[-0.28; 0.44]	1.9%
Boker	11	10.91	3.2100	11	10.55	3.0100		0.11	[-0.73; 0.95]	1.5%
Blomeke, S. et al.	183	93.75	19.0000	183	91.61	17.6000		0.12	[-0.09; 0.32]	2.0%
Luft, J.A., & Zhang, C.	56	4.55	1.3740	56	4.32	1.2077		0.18	[-0.19; 0.55]	1.9%
Luft, J.A., & Zhang, C.	56	3.50	0.8739	56	3.30	1.0255		0.20	[-0.17; 0.58]	1.9%
Buchholtz, N. et al.	59	56.97	11.0600	59	54.31	12.1700		0.23	[-0.13; 0.59]	1.9%
Buchholtz, N. et al.	29	52.67	10.2600	29	50.37	8.0900		0.25	[-0.27; 0.76]	1.8%
Goss, M. et al.	14	65.48	16.7400	14	60.93	14.5700		0.28	[-0.46; 1.03]	1.6%
Blomeke, S. et al.	183	100.00	20.0000	183	93.75	19.0000		0.32	[0.11; 0.53]	2.0%
Castronova	134	3.21	1.5500	134	2.59	1.7400		0.38	[0.13; 0.62]	2.0%
Castronova	134	3.17	1.1900	134	2.64	1.4500		0.40	[0.16; 0.64]	2.0%
Copur-Gencturk, Y. et al.	24	0.55	0.2000	24	0.45	0.2100		0.48	[-0.09; 1.05]	1.7%
Rosenkranzer, F. et al.	23	1.34	0.9500	23	0.91	0.7900		0.48	[-0.10; 1.07]	1.7%
Kulgemeyer, C. et al.	42	0.56	0.1250	47	0.50	0.1210		0.48	[0.06; 0.91]	1.9%
Pape, S.J. et al.	22	0.44	0.8900	23	0.04	0.7100		0.49	[-0.10; 1.08]	1.7%
Rosenkranzer, F. et al.	23	1.39	1.0700	23	0.91	0.7900		0.50	[-0.09; 1.09]	1.7%
Martin, D.A. et al. a	15	0.46	0.0813	15	0.42	0.0775		0.50	[-0.23; 1.23]	1.6%
Rosenkranzer, F. et al.	23	1.00	0.7000	23	0.56	0.9900		0.50	[-0.08; 1.09]	1.7%
Buchholtz, N. et al.	29	50.37	8.0900	29	46.33	7.6800		0.51	[-0.02; 1.03]	1.8%
Luft, J.A., & Zhang, C.	63	5.08	1.4843	63	4.37	1.2611		0.52	[0.16; 0.87]	1.9%
Goss, M. et al.	14	66.22	19.0900	14	56.82	15.4300		0.53	[-0.23; 1.28]	1.5%
Luft, J.A., & Zhang, C.	50	3.82	0.8734	50	3.22	0.9538		0.65	[0.25; 1.05]	1.9%
Smit, R. et al.	121	0.59	0.8700	121	0.02	0.8300		0.67	[0.41; 0.93]	2.0%
Rosenkranzer, F. et al.	23	1.22	0.8500	23	0.56	0.9900		0.70	[0.11; 1.30]	1.7%
Luft, J.A., & Zhang, C.	50	5.06	1.2022	50	4.20	1.1066		0.74	[0.33; 1.14]	1.9%
Martin, D.A. b	17	0.50	0.0880	17	0.43	0.0830		0.79	[0.09; 1.49]	1.6%
Lim-Teo, S.K. et al.	67	34.02	7.4082	67	28.11	7.4082		0.79	[0.44; 1.15]	1.9%
Rosenkranzer, F. et al.	25	1.50	0.8100	25	0.84	0.8000		0.81	[0.23; 1.39]	1.7%
Luft, J.A., & Zhang, C.	63	4.16	1.0350	63	3.30	1.0260		0.83	[0.46; 1.19]	1.9%
Rosenkranzer, F. et al.	25	1.76	0.9300	25	0.84	0.8000		1.04	[0.45; 1.64]	1.7%
Kanter, D.E. et al.	9	3.10	0.7600	9	2.23	0.5700		1.23	[0.20; 2.26]	1.3%
Cheang, W.K. et al.	113	39.83	6.2342	113	32.03	6.2342		1.25	[0.96; 1.53]	2.0%
Kanter, D.E. et al.	9	2.93	0.4300	9	2.15	0.6300		1.38	[0.32; 2.43]	1.2%
Rosenkranzer, F. et al.	23	2.35	0.9100	23	1.12	0.8200		1.40	[0.75; 2.05]	1.7%
Rosenkranzer, F. et al.	23	2.72	0.9700	23	1.26	1.0200		1.44	[0.79; 2.10]	1.6%
Roth, K.J. et al.	71	48.51	4.5500	71	41.98	4.1800		1.49	[1.11; 1.86]	1.9%
Gess-Newsome, J. et al.	27	4.19	1.3000	27	2.24	1.2200		1.52	[0.91; 2.14]	1.7%
Rosenkranzer, F. et al.	23	3.07	1.1200	23	1.41	0.9600		1.56	[0.90; 2.23]	1.6%
Gess-Newsome, J. et al.	27	7.11	1.5500	27	4.52	1.5700		1.64	[1.01; 2.26]	1.7%
Rosenkranzer, F. et al.	23	2.75	1.0100	23	1.12	0.8200		1.74	[1.05; 2.43]	1.6%
Rosenkranzer, F. et al.	23	3.04	0.9800	23	1.26	1.0200		1.75	[1.06; 2.44]	1.6%
Yeo, K.K.J. et al.	113	39.83	4.4160	113	32.03	4.4160		1.76	[1.45; 2.07]	2.0%
Rosenkranzer, F. et al.	25	3.34	0.8700	25	1.44	1.2000		1.78	[1.12; 2.45]	1.6%
Rosenkranzer, F. et al.	23	3.48	1.2900	23	1.41	0.9600		1.79	[1.10; 2.48]	1.6%
Rosenkranzer, F. et al.	23	2.95	1.0100	23	1.19	0.9000		1.81	[1.11; 2.50]	1.6%
Rosenkranzer, F. et al.	25	3.43	0.8100	25	1.70	1.0300		1.84	[1.17; 2.51]	1.6%
Suma, K. et al.	9	74.44	7.2648	9	58.89	7.8174		1.96	[0.79; 3.13]	1.1%
Suma, K. et al.	9	79.11	6.6416	9	61.78	9.8206		1.97	[0.80; 3.14]	1.1%
Rosenkranzer, F. et al.	25	3.53	0.7200	25	1.70	1.0300		2.03	[1.34; 2.72]	1.6%
Rosenkranzer, F. et al.	25	3.70	0.9400	25	1.44	1.2000		2.06	[1.37; 2.76]	1.6%
Suma, K. et al.	9	80.00	8.7464	9	61.00	6.3640		2.37	[1.10; 3.63]	1.1%
Suma, K. et al.	9	77.78	8.2125	9	58.33	7.3655		2.37	[1.10; 3.64]	1.1%
Rosenkranzer, F. et al.	23	3.65	0.8300	23	1.19	0.9000		2.79	[1.96; 3.62]	1.5%
Suma, K. et al.	9	79.11	6.6978	9	57.44	7.0907		2.99	[1.56; 4.43]	0.9%
Random effects model	**2754**			**2760**				**0.90**	**[0.71; 1.10]**	**100.0%**
Prediction interval									**[-0.43; 2.24]**	
Heterogeneity: I^2 = 87%, p < 0.01										

-4   -2   0   2   4

FIGURE 3
Forest plot of effect sizes based on multiple time comparison design.

TABLE 8 Results of the integration of effect sizes of multiple time comparison design based on four models.

Model	k	g [95% CI]	p	95% PI	I^2 [95% CI]
Baseline	59	0.90 [0.71, 1.10]	<0.0001	−0.43, 2.24	87% [85%, 90%]
Trim-and-fill	79	0.44 [0.18, 0.69]	<0.001	−1.66, 2.53	91% [90%, 93%]
Outlier removal	36	0.91 [0.73, 1.09]	<0.0001	0.13, 1.70	63% [47%, 74%]
Three-level model	59/20	0.79 [0.49, 1.09]	<0.0001	−0.56, 2.15	L2 22%, L3 68%

TABLE 9 Results of the meta-regression analysis for the effect sizes of multiple time comparison design (k = 33).

	Estimate	95% CI	t	p
Intercept	0.06	[−0.78, 0.91]	0.16	0.88
USA teachers (In-service teachers)[a]	−0.10	[−0.95, 0.76]	−0.23	0.82
PCK single element (vs. multiple elements)	0.41	[−0.65, 1.48]	0.80	0.43
Science PCK (vs. Math PCK)	1.05	[0.07, 2.02]	2.19	0.04
CK plus PCK Intervention (vs. CK only)	0.21	[−0.31, 0.73]	0.82	0.42
Longer intervention (vs. shorter intervention)	−0.46	[−1.79, 0.87]	−0.71	0.48

[a]In the data for this study, the variables of teachers' countries (USA vs. German teachers) and teacher status (pre- vs. in-service teachers) are perfectly confounded, in which, all USA teachers were in-service teachers, while all German teachers were pre-service teachers. As a result, these two factors were completely confounded in this study.

mathematics. Hands-on activities, whereby teachers themselves conduct scientific experiments or computer simulations, are used more frequently in professional development for science education than for mathematics education. Several studies have actually introduced these activities to enhance teachers' PCK (Kanter and Konstantopoulos, 2010; Rosenkränzer et al., 2017; Yang et al., 2018). However, various other differences exist between PCK studies in mathematics and science. In studies focusing on mathematics, for example, tests often cover a wide range of topics, such as functions and quantities, rather than specific contents (e.g., Copur-Gencturk and Lubienski, 2013). In contrast, science studies often focus on specific domains such as biology or chemistry, and within those domains, some of the studies address specific topics like photosynthesis (e.g., Gess-Newsome et al., 2019). This difference may lead to more pronounced effects of the interventions in science rather than mathematics.

Finally, we discuss the remaining issues of this study. First, it is necessary to clarify the effects of interventions that were not addressed in the meta-regression analysis of this study, such as the characteristics of the intervention method. Even if the interventions have the same elements and duration, their effects are expected to vary depending on the method used in the intervention. For example, in Roth et al. (2019), one feature of the intervention in the experimental group was a video recording of a model lesson or their own lesson and an activity to watch it. However, this study has not been able to examine the effects of these activities. It has been noted that even in the whole area of professional development research of teachers, demonstrating the effectiveness of intervention methods has not always been successful (Garet et al., 2001; Hill et al., 2013). Although we were unable to examine the effects of methods in the studies included in this study because the number of effect sizes were limited, future studies should clarify the effects of characteristics in the intervention methods on the acquisition of PCK.

Second, it is assumed that not only the characteristics of the intervention but also the characteristics of the evaluation (how PCK is assessed) might affect the effect size of the intervention on PCK (e.g., Copur-Gencturk and Lubienski, 2013). In addition to the elements of PCK, the present study also reported the proportion of PCK formats (multiple-choice, open-ended, and other formats). Since it is assumed that the degree to which teachers themselves have to utilize their knowledge is great in the open-ended format (e.g., Fukaya and Uesaka, 2018), the effect of the intervention could be evaluated more clearly in the non-multiple-choice format. However, we were not able to examine such effects in this study, so further studies are expected in the future.

In addition, the transfer of learning needs to be examined as an perspective that has not been focused on in previous studies (for literature on the transfer of learning, see Bransford et al., 1999; Haskell, 2001; Mestre, 2005). Transfer of learning goes beyond simply remembering what has been learned and refers to applying what has been learned in a new task. In teacher education, teachers are expected to apply the knowledge and skills acquired in training and university to content not covered in their classroom. Therefore, what is expected in PCK interventions is not just recall but transfer of what they learn in the interventions. It would be important to distinguish in intervention studies whether the index used for evaluation is at the mere memory level or at the transfer level (i.e., whether it includes content not covered in the intervention). As revealed in this study, few studies to date have explicitly explained how the measures used in the evaluation relate to the intervention, so we could not determine whether the evaluation index was at the transfer level or not. It seems essential for future studies to specify the key features of the intervention or evaluation in the paper, including whether they evaluated the transfer of learning.

Author's note

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

TF: Conceptualization, Data curation, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing. DN: Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. YK: Data curation, Investigation, Writing – original draft, Writing – review & editing. TN: Data curation, Investigation, Writing – original draft, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abell, S. K. (2008). Twenty years later: does pedagogical content knowledge remain a useful idea? *Int. J. Sci. Educ.* 30, 1405–1416. doi: 10.1080/09500690802187041

Alonzo, A. C., and Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *J. Res. Sci. Teach.* 53, 1259–1286. doi: 10.1002/tea.21271

Ball, D. L., Lubienski, S. T., and Mewborn, D. S. (2001). "Research on teaching mathematics: the unsolved problem of teachers' mathematical knowledge," in *Handbook of research on teaching,* ed. V. Richardson (Washington, DC: American Educational Research Association), 433–456.

Ball, D. L., Thames, M. H., and Phelps, G. (2008). Content knowledge for teaching: what makes it special? *J. Teach. Educ.* 59, 389–407. doi: 10.1177/0022487108324554

Baumert, J., and Kunter, M. (2013). "The COACTIV model of teachers' professional competence," in *Cognitive activation in the mathematics classroom and professional competence of teachers,* eds. M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand (Cham: Springer), 25–48. doi: 10.1007/978-1-4614-5149-5_2

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. *Am. Educ. Res. J.* 47, 133–180. doi: 10.3102/0002831209345157

Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S., and Neubrand, M. (2013). "Professional competence of teachers, cognitively activating instruction, and the development of students mathematical literacy (COACTIV): a research program," in *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project,* eds. M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand (Cham: Springer), 1–21. doi: 10.1007/978-1-4614-5149-5_1

Becker, B. J. (2000). "Multivariate meta-analysis," in *Handbook of applied multivariate statistics and mathematical modeling,* eds. H. E. A. Tinsley and E. D. Brown (Orlando, FL: Academic Press), 499–525. doi: 10.1016/B978-012691360-6/50018-5

Bednarz, N., and Proulx, J. (2009). Knowing and using mathematics in teaching: conceptual and epistemological clarifications. *Learn. Mathem.* 29, 11–17.

*Besser, M., Leiss, D., and Blum, W. (2020). Who participates in which type of teacher professional development? Identifying and describing clusters of teachers. *Teach. Dev.* 24, 293–314. doi: 10.1080/13664530.2020.1761872

*Blömeke, S., Buchholtz, N., Suhl, U., and Kaiser, G. (2014). Resolving the chicken-or-egg causality dilemma: the longitudinal interplay of teacher knowledge and teacher beliefs. *Teach. Teach. Educ.* 37, 130–139. doi: 10.1016/j.tate.2013.10.007

Blömeke, S., and Kaiser, G. (2014). "Theoretical framework, study design and main results of TEDS-M," in *International perspectives on teacher knowledge, beliefs and opportunities to learn. TEDS-M results,* eds. S. Blömeke, F.-J. Hsieh, G. Kaiser, and W. H. Schmidt (Cham: Springer), 19–47. doi: 10.1007/978-94-007-6437-8_2

*Boker, J. R. (2016). *Developing Design Expertise Through a Teacher-Scientist Partnership Professional Development Program.* [Unpublished doctoral dissertation]. The University of Florida.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to Meta-Analysis.* Hoboken, NJ: Wiley. doi: 10.1002/9780470743386

Bransford, J. D., Brown, A. L., and Cocking, R. R. (1999). *How People Learn: Brain, Mind, Experience, and School.* Washington DC: National Academy Press

Brophy, J., and Good, T. (1986). Teacher behavior and student achievement," in *Third handbook of research on teaching,* ed. M. Wittrock (New York: Macmillan), 328–375.

Brownlee, J., Purdie, N., and Boulton-Lewis, G. (2001). Changing epistemological beliefs in pre-service teacher education students. *Teach. Higher Educ.* 6, 247–268. doi: 10.1080/13562510120045221

*Buchholtz, N., and Kaiser, G. (2013). Improving mathematics teacher education in Germany: empirical results from a longitudinal evaluation of innovative programs. *Int. J. Sci. Mathem. Educ.* 11, 949–977. doi: 10.1007/s10763-013-9427-7

Callingham, R., Carmichael, C., and Watson, J. M. (2016). Explaining student achievement: The influence of teachers' pedagogical content knowledge in statistics. *Int. J. Sci. Mathem. Educ.* 14, 1339–1357. doi: 10.1007/s10763-015-9653-2

*Castronova, M. A. (2018). *Examining teachers' acceptance of the next generation science standards: a study of teachers' pedagogical discontentment and pedagogical content knowledge of modeling and argumentation* [Unpublished doctoral dissertation]. Caldwell University.

Chan, K. K. H., and Hume, A. (2019). "Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated," in *Repositioning PCK in teachers' professional knowledge,* eds. A. Hume, R. Cooper, and A. Borowski (Cham: Springer), 3–76. doi: 10.1007/978-981-13-5898-2_1

*Cheang, W. K., Yeo, J. K. K., Chan, E. C. M., and Lim-Teo, S. K. (2007). Development of mathematics pedagogical content knowledge in student teachers. *Mathem. Educ.* 10, 27–54.

Cochran, K. F., DeRuiter, J. A., and andKing, R. A. (1993). Pedagogical content knowing: an integrative model for teacher preparation. *J. Teach. Educ.* 44, 263–272. doi: 10.1177/0022487193044004004

*Copur-Gencturk, Y., and Lubienski, S. T. (2013). Measuring mathematical knowledge for teaching: a longitudinal study using two measures. *J. Mathem. Teach. Educ.* 16, 211–236. doi: 10.1007/s10857-012-9233-0

Dalgarno, N., and Colgan, L. (2007). Supporting novice elementary mathematics teachers' induction in professional communities and providing innovative forms of pedagogical content knowledge development through information and communication technology. *Teach. Teach. Educ.* 23, 1051–1065. doi: 10.1016/j.tate.2006.04.037

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Educ. Policy Anal. Arch.* 8, 1–44. doi: 10.14507/epaa.v8n1.2000

Depaepe, F., Verschaffel, L., and Kelchtermans, G. (2013). Pedagogical content knowledge: a systematic review of the way in which the concept has pervaded mathematics educational research. *Teach. Teach. Educ.* 34, 12–25. doi: 10.1016/j.tate.2013.03.001

Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Rev. Res. Educ.* 5, 163–198. doi: 10.2307/1167174

Dreher, A., and Kuntze, S. (2015). Teachers' professional knowledge and noticing: the case of multiple representations in the mathematics classroom. *Educ. Stud. Mathem.* 88, 89–114. doi: 10.1007/s10649-014-9577-8

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315, 629–634. doi: 10.1136/bmj.315.7109.629

Evens, M., Elen, J., and Depaepe, F. (2015). Developing pedagogical content knowledge: lessons learned from intervention studies. *Educ. Res. Int.* 2015, 1–23. doi: 10.1155/2015/790417

Evens, M., Elen, J., Larmuseau, C., and Depaepe, F. (2018). Promoting the development of teacher professional knowledge: integrating content and pedagogy in teacher education. *Teach. Teach. Educ.* 75, 244–258. doi: 10.1016/j.tate.2018.07.001

Fauth, B., Decristan, J., Decker, A. T., Büttner, G., Hardy, I., Klieme, E., et al. (2019). The effects of teacher competence on student outcomes in elementary science education: the mediating role of teaching quality. *Teach. Teach. Educ.* 86:102882. doi: 10.1016/j.tate.2019.102882

Fennema, E., and Franke, M. L. (1992). Teachers' knowledge and its impact," in *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics,* ed. D. A. Grouws (New York: Macmillan Publishing Co, Inc), 147–164.

Förtsch, S., Förtsch, C., Von Kotzebue, L., and Neuhaus, B. J. (2018). Effects of teachers' professional knowledge and their use of three-dimensional physical models in biology lessons on students' achievement. *Educ. Sci.* 8, 1–28. doi: 10.3390/educsci8030118

Fukaya, T., and Uesaka, Y. (2018). Using a tutoring scenario to assess the spontaneous use of knowledge for teaching. *J. Educ. Teach.* 44, 431–445. doi: 10.1080/02607476.2018.1450821

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., and Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *Am. Educ. Res. J.* 38, 915–945. doi: 10.3102/00028312038004915

Gess-Newsome, J. (2013). "Pedagogical content knowledge," in *International Guide to Student Achievement,* eds. J. Hattie and E. M. Anderman (New York, NY: Routledge), 257–259.

*Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D., and Stuhlsatz, M. A. (2019). Teacher pedagogical content knowledge, practice, and student achievement. *Int. J. Sci. Educ.* 41, 944–963. doi: 10.1080/09500693.2016.1265158

*Goss, M., Powers, R., and Hauk, S. (2013). Identifying change in secondary mathematics teachers' pedagogical content knowledge," in *Proceedings for the 16th conference on Research in Undergraduate Mathematics Education* (Denver: SIGMAA).

Großschedl, J., Harms, U., Kleickmann, T., and Glowinski, I. (2015). Preservice biology teachers' professional knowledge: structure and learning opportunities. *J. Sci. Teacher Educ.* 26, 291–318. doi: 10.1007/s10972-015-9423-6

Harr, N., Eichler, A., and Renkl, A. (2014). Integrating pedagogical content knowledge and pedagogical/psychological knowledge in mathematics. *Front. Psychol.* 5:96118. doi: 10.3389/fpsyg.2014.00924

Harr, N., Eichler, A., and Renkl, A. (2015). Integrated learning: ways of fostering the applicability of teachers' pedagogical and psychological knowledge. *Front. Psychol.* 6:124369. doi: 10.3389/fpsyg.2015.00738

Harrer, M., Cuijpers, P., Furukawa, T. A., and Ebert, D. D. (2021). *Doing Meta-Analysis with R: A Hands-on Guide.* New York, NY: Chapman and Hall/CRC Press. doi: 10.1201/9781003107347

Haskell, R. E. (2001). *Transfer of Learning: Cognition, Instruction, and Reasoning.* San Diego, CA: Academic Press.

Haston, W., and Leon-Guerrero, A. (2008). Sources of pedagogical content knowledge: reports by preservice instrumental music teachers. *J. Music Teach. Educ.* 17, 48–59. doi: 10.1177/1057083708317644

Hattie, J. A. C. (2009). *Visible Learning: A Synthesis of 800 Meta-Analyses on Achievement.* London: Routledge.

Higgins, J., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *Br. Med. J.* 327, 557–560. doi: 10.1136/bmj.327.7414.557

Hill, H. C., Ball, D. L., and Schilling, S. G. (2008). Unpacking pedagogical content knowledge: conceptualizing and measuring teachers' topic-specific knowledge of students. *J. Res. Mathem. Educ.* 39, 372–400. doi: 10.5951/jresematheduc.39.4.0372

Hill, H. C., Beisiegel, M., and Jacob, R. (2013). Professional development research: consensus, crossroads, and challenges. *Educ. Resear.* 42, 476–487. doi: 10.3102/0013189X13512674

Hill, H. C., Rowan, B., and Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *Am. Educ. Res. J.* 42, 371–406. doi: 10.3102/00028312042002371

Hill, H. C., Schilling, S. G., and Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Element. School J.* 105, 11–30. doi: 10.1086/428763

Hogan, T., Rabinowitz, M., and Craven, I. I. I., J. A. (2003). Representation in teaching: inferences from research of expert and novice teachers. *Educ. Psychol.* 38, 235–247. doi: 10.1207/S15326985EP3804_3

Kanter, D. E., and Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: the effects of teacher content and pedagogical content knowledge and inquiry-based practices. *Sci. Educ.* 94, 855–887. doi: 10.1002/sce.20391

Keller, M. M., Neumann, K., and Fischer, H. E. (2016). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *J. Res. Sci. Teach.* 54, 586–614. doi: 10.1002/tea.21378

Kenney, R., Shoffner, M., and Norris, D. (2013). Reflecting to learn mathematics: Supporting pre-service teachers' pedagogical content knowledge with reflection on writing prompts in mathematics education. *Reflect. Pract.* 14, 787–800. doi: 10.1080/14623943.2013.836082

Kind, V. (2009). Pedagogical content knowledge in science education: perspectives and potential for progress. *Stud. Sci. Educ.* 45, 169–204. doi: 10.1080/03057260903142285

Knapp, G., and Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* 22, 2693–2710. doi: 10.1002/sim.1482

*Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., et al. (2020). Professional knowledge affects action-related skills: the development of preservice physics teachers' explaining skills during a field experience. *J. Res. Sci. Teach.* 57, 1554–1582. doi: 10.1002/tea.21632

*Lenhart, S. T. (2010). *The effect of teacher pedagogical content knowledge and the instruction of middle school geometry* [Unpublished doctoral dissertation]. Liberty University.

*Lim-Teo, S. K., Chua, K. G., Cheang, W. K., and Yeo, J. K. (2007). The development of diploma in education student teachers' mathematics pedagogical content knowledge. *Int. J. Sci. Mathem. Educ.* 5, 237–261. doi: 10.1007/s10763-006-9056-5

*Luft, J. A., and Zhang, C. (2014). The pedagogical content knowledge and beliefs of newly hired secondary science teachers: the first three years. *Educación Química* 25, 325–331. doi: 10.1016/S0187-893X(14)70548-8

*Martin, D. A. (2017). *The impact of problem-based learning on pre-service teachers' development and application of their mathematics pedagogical content knowledge.* Doctoral dissertation (University of Southern Queensland).

*Martin, D. A., Grimbeek, P., and Jamieson-Proctor, R. (2013). "Measuring problem-based learning's impact on pre-service teachers' mathematics pedagogical content knowledge," in *Proceedings of the 2nd International Higher Education Teaching and Learning Conference (IEAA 2013)* (University of Southern Queensland).

Mason, J. (2008). "PCK and beyond," in *International handbook of mathematics teacher education: Vol. 1. Knowledge and beliefs in mathematics teaching and teaching development,* eds. P. Sullivan, T. Wood (Dordrecht: Sense Publishers), 301–322.

*Matthews, M. E. (2006). *The effects of a mathematics course designed for preservice elementary teachers on mathematical knowledge and attitudes.* [Unpublished doctoral dissertation]. The University of Iowa.

*Mavhunga, M. E. (2012). *Explicit inclusion of topic specific knowledge for teaching and the development of PCK in pre-service science teachers* [Unpublished doctoral dissertation]. University of the Witwatersrand.

Meschede, N., Fiebranz, A., Möller, K., and Steffensky, M. (2017). Teachers' professional vision, pedagogical content knowledge and beliefs: on its relation and differences between pre-service and in-service teachers. *Teach. Teach. Educ.* 66, 158–170. doi: 10.1016/j.tate.2017.04.010

Mestre, J. P. (2005). *Transfer of Learning From a Modern Multidisciplinary Perspective*. Charlotte, NC: IAP.

Oh, P. S., and Kim, K. S. (2013). Pedagogical transformations of science content knowledge in Korean elementary classrooms. *Int. J. Sci. Educ.* 35, 1590–1624. doi: 10.1080/09500693.2012.719246

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst. Rev.* 10, 1–11. doi: 10.1186/s13643-021-01626-4

*Pape, S. J., Prosser, S. K., Griffin, C. C., Dana, N. F., Algina, J., and Bae, J. (2015). Prime online: developing grades 3-5 teachers' content knowledge for teaching mathematics in an online professional development program. *Contemp. Issues Technol. Teach. Educ.* 15, 14–43.

Petrou, M., and Goulding, M. (2011). "Conceptualising teachers' mathematical knowledge in teaching," in *Mathematical knowledge in teaching*, eds. T. Rowland, K. Ruthven (Cham: Springer), 9–25. doi: 10.1007/978-90-481-9766-8_2

Pigott, T. D., and Polanin, J. R. (2020). Methodological guidance paper: high-quality meta-analysis in a systematic review. *Rev. Educ. Res.* 90, 24–46. doi: 10.3102/0034654319877153

Polanin, J. R., Hennessy, E. A., and Tsuji, S. (2020). Transparency and reproducibility of meta-analyses in psychology: a meta-review. *Persp. Psychol. Sci.* 15, 1026–1041. doi: 10.1177/1745691620906416

*Rosenkränzer, F., Hörsch, C., Schuler, S., and Riess, W. (2017). Student teachers' pedagogical content knowledge for teaching systems thinking: effects of different interventions. *Int. J. Sci. Educ.* 39, 1932–1951. doi: 10.1080/09500693.2017.1362603

Rosenshine, B. (1976). Recent research on teaching behaviors and student achievement. *J. Teach. Educ.* 27, 61–64. doi: 10.1177/002248717602700115

*Roth, K. J., Wilson, C. D., Taylor, J. A., Stuhlsatz, M. A., and Hvidsten, C. (2019). Comparing the effects of analysis-of-practice and content-based professional development on teacher and student outcomes in science. *Am. Educ. Res. J.* 56, 1217–1253. doi: 10.3102/0002831218814759

Seidel, T., and Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Rev. Educ. Res.* 77, 454–499. doi: 10.3102/0034654307310317

Seymour, J. R., and Lehrer, R. (2006). Tracing the evolution of pedagogical content knowledge as the development of interanimated discourses. *J. Learn. Sci.* 15, 549–582. doi: 10.1207/s15327809jls1504_5

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educ. Resear.* 15, 4–14. doi: 10.2307/1175860

Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harv. Educ. Rev.* 57, 1–23. doi: 10.17763/haer.57.1.j463w79r56455411

Simmons, P. E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., et al. (1999). Beginning teachers: beliefs and classroom actions. *J. Res. Sci. Teach.* 36, 930–954. doi: 10.1002/(SICI)1098-2736(199910)36:8<930::AID-TEA3>3.0.CO;2-N

*Smit, R., Weitzel, H., Blank, R., Rietz, F., Tardent, J., and Robin, N. (2017). Interplay of secondary pre-service teacher content knowledge (CK), pedagogical content knowledge (PCK) and attitudes regarding scientific inquiry teaching within teacher training. *Res. Sci. Technol. Educ.* 35, 477–499. doi: 10.1080/02635143.2017.1353962

Sorge, S., Kröger, J., Petersen, S., and Neumann, K. (2019). Structure and development of pre-service physics teachers' professional knowledge. *Int. J. Sci. Educ.* 41, 862–889. doi: 10.1080/09500693.2017.1346326

*Strawhecker, J. (2005). Preparing elementary teachers to teach mathematics: how field experiences impact pedagogical content knowledge. *Issues Undergr. Mathem. Prepar. School Teach.* 4, 1–12.

*Suma, K., Sadia, I. W., and Pujani, N. M. (2019). The effect of lesson study on science teachers' pedagogical content knowledge and self-efficacy. *Int. J. New Trends Educ. Their Implic.* 10, 1–5.

Tröbst, S., Kleickmann, T., Heinze, A., Bernholt, A., Rink, R., and Kunter, M. (2018). Teacher knowledge experiment: testing mechanisms underlying the formation of preservice elementary school teachers' pedagogical content knowledge concerning fractions and fractional arithmetic. *J. Educ. Psychol.* 110, 1049–1065. doi: 10.1037/edu0000260

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J. Educ. Behav. Statist.* 30, 261–293. doi: 10.3102/10769986030003261

Viechtbauer, W., and Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Res. Synth. Methods* 1, 112–125. doi: 10.1002/jrsm.11

Wayne, A. J., and Youngs, P. (2003). Teacher characteristics and student achievement gains: a review. *Rev. Educ. Res.* 73, 89–122. doi: 10.3102/00346543073001089

*Yang, Y., Liu, X., and Gardella Jr, J. A. (2018). Effects of professional development on teacher pedagogical content knowledge, inquiry teaching practices, and student understanding of interdisciplinary science. *J. Sci. Teacher Educ.* 29, 263–282. doi: 10.1080/1046560X.2018.1439262

*Yeo, K. K. J., Cheang, W. K., and Chan, C. M. E. (2006). "Development of mathematics pedagogical content knowledge in pre-service teachers," in *Proceedings of APERA Conference*, 1–10.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., and Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement (Issues and Answers Report, REL 2007—No. 033)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.