Professional Development Interventions for Mathematics Teachers: A Systematic Review.

Branko Bognar[1], Ljerka Jukić Matić[2], Marija Sablić[1]

[1]Faculty of Humanities and social Sciences, Department of Pedagogy, University of Osijek, Osijek, Croatia,

[2]School of Applied Mathematics and Informatics, University of Osijek, Osijek, Croatia

branko.bognar@gmail.com

Abstract: In many countries around the world, stakeholders engaged in driving education reform policy use teacher professional development to improve the quality of teacher learning, expecting a positive effect on the quality of teaching. Given the high level of expectations for professional development, it is crucial to identify the characteristics of effective teacher professional development. Therefore, we conducted a systematic literature review of professional development interventions for mathematics teachers. We sought to identify the characteristics of interventions with positive and statistically significant effects on students' mathematics achievement. Our review includes 12 professional development interventions which included elements of structured pedagogy intervention (i.e., teacher training, on-site teacher support, and resources for teachers and students), in addition to initial professional development and follow-up workshops. Utilizing technology has proved to be beneficial for student learning, but less so for teacher learning. The results of the reviewed studies indicate that changes in instruction can be implemented incrementally, beginning with less complex interventions and progressing to those that are more complex and demanding. Furthermore, we conclude that professional development interventions that seek to improve student learning outcomes in mathematics should include on-site teacher support, mentoring and feedback, teacher-focused resources, and classroom learning materials.

Keywords: effective mathematics interventions, mathematics education, systematic review, professional development

INTRODUCTION

In many countries, education reform policymakers use teacher professional development to improve teacher learning, which is expected to improve teaching, student achievement, and teachers' long-term beliefs and attitudes. Professional development (PD) consists of unplanned and planned learning experiences and activities that enhance teachers' knowledge, attitudes, and skills, as well as their teaching practices (Day & Leitch, 2007; Avalos, 2011). Teacher professional development is defined as learning that can be either individual or collective, but should be contextualized in the teacher's workplace — the school — and contributes to the development of competencies through a variety of formal and informal experiences" (Marcelo, 2009). Professional development includes “the body of systematic activities [used] to prepare teachers for their job, including initial training, induction training, in-service training, and continuous professional development within school settings” (Hendriks et al., 2010, p. 19). Consequently, PD includes both pre-service and in-service teacher education (Bautista & Oretga-Ruiz, 2017). Guskey (2002) notes that contemporary teachers expect in-service training to provide them with concrete and specific procedures they can implement daily in the classroom.

Fullan and Hargreaves (2016) distinguish between professional development and professional learning, although some authors use the terms interchangeably. According to them, professional development consists of activities for their own sake, whereas professional learning is characterized by measurable quality, performance, and teacher impact. According to Fraser et al. (2007), professional development ensures teachers' professionalism, while professional learning results in “specific changes in the professional knowledge, skills, attitudes, beliefs, or actions of teachers.” However, Evans (2014) notes that professional development is the process by which people's professionalism can be viewed as being enhanced with a degree of permanence that exceeds transience. This paper will use the term professional development to mean planned, collective, in-service learning activities with the core aim of improving teachers’ competencies and teaching practices, and which strive to contribute to the quality of student learning.

In mathematics teacher professional development, it is vital to recognize the various types of knowledge. These include content knowledge (CK), which is mathematical knowledge, pedagogical content knowledge (PCK), which is “ways of representing and formulating the subject that make it comprehensible to others” (Shulman, 1986, p. 9), and pedagogical knowledge (PK), which is “independent knowledge of how to optimise learning situations in the classroom in general” (Krauss et al., 2008, p. 874). Ball et al. (2008) separate mathematical knowledge from pedagogical knowledge. Pedagogical knowledge encompasses content, students, instruction, and curriculum. Therefore, pedagogical content knowledge necessitates an understanding of students' mathematical reasoning, their interests, and their areas of difficulty. This latter type of knowledge enables teachers to design and implement lesson plans in accordance with the subject curriculum, taking student needs into account, using a variety of teaching methods, mathematical concept presentation methods, and suitable examples.

Given the high level of interconnected expectations currently placed on professional development, it is crucial to identify the characteristics of effective teacher professional development. Specifically, we are interested in factors that have positive and statistically significant effects on students' mathematical achievement. In order to rigorously investigate this issue, we conducted a systematic literature review focusing on randomized experimental studies conducted in the past 21 years (i.e., 1999–2020) in order to identify the characteristics of effective interventions in terms of students' mathematical achievement. Our analysis encompasses not only professional development programs but also multicomponent interventions that incorporate a variety of instructional and pedagogical techniques. Our decision to include multicomponent interventions comes from Hull et al. (2018), who contend that multicomponent interventions are particularly pertinent in real-world applications that seek to improve educational outcomes, such as student achievement. Therefore, we were interested in any randomized controlled trial (RCT) that evaluated instances in which teacher professional development led to statistically significant gains in student mathematics achievement. By conducting this systematic review, we tried to answer the following research question: What are the features of professional development interventions for mathematics teachers that have a statistically significant and positive effect size?

METHODS

The process of conducting systematic review began in the middle of 2021; therefore, we decided to include available randomized controlled trials published in English language between 1999 and 2020, carried out across all grades of elementary, middle, and high schools that include PD programs for mathematics teachers. We have chosen 1999 as our starting point because of Kennedy (1998), who conducted a systematic review of professional development programmes aimed at enhancing student learning in mathematics and science. The method used in this review is based on the emerging literature addressing this multicomponent approach (i.e., Petticrew & Roberts, 2006; Gough et al., 2017; Polanin et al., 2017; Siddaway et al., 2019), a procedure consisting of several steps. First, we formulated the research question. Second, we defined the search terms and selected appropriate databases. Third, we used inclusion and exclusion criteria to ensure the scientific quality of the relevant publications (Table 1). Finally, the data answering the research question was extracted, analysed, and interpreted.

For quantitative studies, the effect size (i.e., related to students' mathematics achievement) and its statistical significance (p value) are especially important. To assess its statistical significance, the effect size should be at least conventionally significant ($p <.05$). However, there is no clear criteria by which to determine whether the effect size is practically significant as it depends on numerous diverse factors such as sample size; research design; measure type; and differences among students, teaching subjects, schools, etc. (Hill et al., 2008; Slavin & Smith, 2009; Lipsey et al., 2012; Cheung & Slavin, 2016; Kraft, 2020), thus could hardly be unambiguously interpreted. In a bid to seek some clarity and consistency in this regard, Bakker et al. (2019) suggest 12 points for interpreting effect sizes in mathematics education journals some of which are technical (e.g. Which calculation of the effect size is used? or What is the confidence interval around the point estimate?), some methodological (e.g. To what is the effect compared? or Focus on offering or receiving?), an some empirical/ontological (e.g. What is the context? or What was the sample?). In summary, those points cover research design, alignment of the intervention and measurement, intervention duration, sample size, and context. Bearing in mind this potential for ambiguity and multiplicity, we have included various key factors in Table 2 (i.e., country where the study was conducted, study duration, sample size, and type of measure) in order to allow for different interpretations of intervention impacts. For example, if an education system is of a low standard, it may have more room for improvement (Bakker et al., 2019), meaning that interventions in such a context may result in higher effect sizes. Similarly, an intervention carried out over a longer period of time may produce a larger effect size (Bakker et al., 2019), as might a program in which the researchers develop their own outcome measure (Lipsey et al., 2012). In contrast, however, Slavin & Smith (2009) found a statistically negative correlation between effect size and sample size in their research on elementary and secondary mathematics programs.

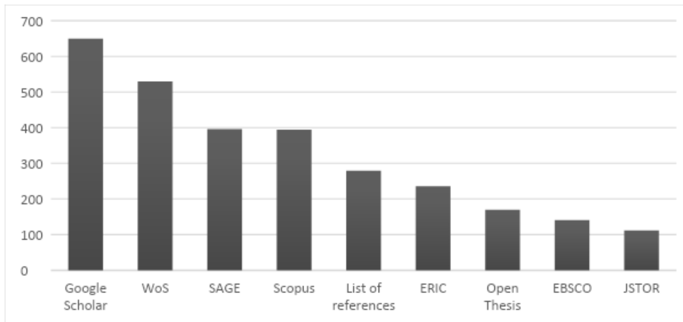

Figure 1. Number of studies included in the analysis, arranged by source.

Our main resources for finding relevant literature — empirical studies published between 1999 and 2020 — were electronic databases. We focused on peer-reviewed journal articles and non-academic research publications (commonly called grey literature). Our search for relevant publications included these databases: EBSCO, Education Resource Center (ERIC), Google Scholar, JSTOR, SAGE, SCOPUS, and Web of Science (see Figure 1). We used the following search terms in the titles, abstracts, and keywords of articles: "professional development" OR "professional learning" OR "in-service education" OR "in-service training", AND math or science AND experiment OR trial. The use of these search terms, which are in line with prior discussions in this paper, was intended to limit the data collection on the sources connected to our research questions. Due to the fact that we were conducting a systematic review of the literature regarding the professional development of science teachers at the same time (published in a separate publication), we included the keyword "science" in the search. In the subsequent analysis, the research related to science was separated from that related to the professional development of mathematics teachers. The use of above search terms was intended to limit the data collection to sources connected to our research question. In order to identify relevant grey literature for the review (e.g., reports, academic theses, working papers, etc.), we searched relevant targeted online repositories, such as Google Scholar and Open Thesis. In addition, we examined lists of references in selected articles, systematic reviews, and meta-analyses related to the professional development of mathematics teachers. As a result, we yielded 2,908 potential publications (i.e. academic papers and grey literature publications) featuring a relationship between our keywords.

Next, we carried out a preliminary review of these 2,908 publications, searching for studies that reported only on randomized design comparing treatment groups with groups using existing programs, plus that included PD focused on mathematics and students' mathematics achievements (reported either as a calculable effect size or quantitative results). We also followed Slavin's guidelines whereby selected studies should include a pre-test stage, although "randomized experiments without pre-tests are acceptable if attrition is low and equal between experimental and control groups" (2008, p. 8). Furthermore, we employed the guideline that treatment and control groups must include at least two teachers and 30 students (Cheung and Slavin, 2016), where a combination of differential and overall attrition is within an 'optimistic' boundary (What Works Clearinghouse, 2020). Lastly, we required that an intervention lasted at least 12 weeks (Pellegrini et al., 2018) and that the PD program was described in detail. Inclusion criteria can be seen in Table 1. We excluded non-randomized experiments, evaluation research, and randomized designs that did not examine students' mathematics achievement. We utilized Rayyan QCRI, a free web application designed specifically for systematic reviews and other knowledge synthesis initiatives (Ouzzani et al., 2016) for this section of the systematic review. The utilization of this application aided in the process of screening and selection of studies. The initial screening yielded 54 documents that were further examined according to the inclusion and exclusion criteria (Table 1).

Inclusion criteria	Exclusion criteria
Students in grades 1-12	Pre-kindergarten, kindergarten children as well as students in postsecondary education
In-service professional development for mathematics teachers was part of the intervention	Professional development for mathematics teacher is related to pre-service teacher education
The professional development programme is described in detail	There is insufficient information regarding the professional development programme
Randomised experiment comparing treatment groups with groups using business-as-usual or other programme already in place	Non-randomized experiment, non-experimental research, or randomized experiment without a control group
Study comprises pre-test although "randomized experiments without pre-tests are acceptable if attrition is low and equal between experimental and control groups" (Slavin, 2008, p. 8)	Randomized experiments without pre-test and with high attrition between groups along with those "in which pre-test differences are more than 50% of a standard deviation" (Slavin, 2008, p. 8)
Studies with at least two teachers and 30 students in treatment and control groups (Cheung & Slavin, 2016, p. 286)	Either treatment or control group have only one teacher
Study includes quantitative measures of students' mathematics outcomes (e.g. standardized test or a test that was developed by researcher which is fair to all treatment and control groups)	Study includes only qualitative data, or quantitative measures without students' mathematics performance, or dependant measures favour some of the treatment groups

Inclusion criteria	Exclusion criteria
Effect size for students' mathematics outcomes is calculated using appropriate analysis or it is possible to calculate it from given results	Effect size is not included, or it is calculated by inappropriate analysis and it is not possible to calculate/correct it from given results
Study had positive effect sizes.	Study had negative or statistically insignificant effect sizes.
Interventions lasting at least 12 weeks "to make it more likely that effective programmes could be replicated over extended periods" (Pellegrini et al., 2018, p. 8)	Interventions lasting less than 12 weeks
Study was published from 1999 to 2020	Study was published before 1999 or after 2020
The research could be conducted in any country, but the paper must be written in English	The paper is not written in English

Table 1: Inclusion and exclusion criteria

The inclusion and exclusion criteria (Table 1) were applied to the full-text versions of the remaining articles. This yielded 12 publications that we selected for our systematic literature review (see Table 2). For some of these selected interventions, we found additional studies that gave us better insight into PD programs. Most of the selected studies were conducted in the USA (8), and there was one study each from the following countries: Belize, China, Canada, and Pakistan. In terms of the year of publication, most studies (8) were published recently, i.e. between 2016 and 2020. Although our intention was to include studies published in the last 21 years, we did not find any published before 2007 that met the inclusion criteria. While reading the full-text versions of the selected publications, we extracted the relevant data necessary for answering the research question. For this part of the research, we used EPPI Reviewer Web (https://eppi.ioe.ac.uk/EPPIReviewer-Web), a web-based software designed for various types of literature review, including systematic reviews. This extraction of data was done using categories that we determined by analysing the selected PD programs (i.e., initial professional development, follow-up workshops, coaching, online learning, use of videos, and types of teacher knowledge).

RESULTS AND DISCUSSION

In our analysis of the included studies, we first focused on changes in teaching, because without changes in teaching it is difficult to expect any improvement in student results (Guskey, 2002; Kunter et al., 2013; Campbell et al., 2014; Kennedy, 2016). All analysed interventions have resulted in positive changes to teaching practices, student knowledge, and their mathematics achievement. Within this, we identified various teaching strategies which positioned students as active participants in the teaching-learning process: i.e., where efforts were made to increase student engagement in line with standards and curricula, providing them with optimal challenges to reach a higher level than they were previously at (Early et al., 2016). As such, teachers elicited deeper levels of student thinking, reasoning, higher-order thinking skills, and learning (Newman et al., 2012; Lewis & Perry, 2017; Chen et., al. 2020).

Title of program (Reference)	Country / Study duration / School year(s)	Aim	Baseline sample size	Type of measure, effect size†
1. eLearn (Beg et al. 2019)	Pakistan / 4 months / 2016/2017	Examining the effectiveness of short videos on student achievement in mathematics and science	60 schools, 274 eighth-grade teachers, 2,999 students	c) .26
2. Video-based teacher professional development (Chen et al., 2020)	China / 1 year / n/a	Investigating the efficacy of video-based professional development programs using a discourse visualization tool	16 schools, 54 sixth- and seventh-grade teachers, 1,507 students	c) .24
3. Every Classroom, Every Day (Early et al., 2016)	USA / 2 years / 2009/2010, 2010/2011	Instructional intervention aimed at increasing students' learning and achievement	20 high schools, n/a ninth- and tenth-grade teachers, 8,250 students	a) .15*
4. Math Pathways and Pitfalls (Heller et al., 2007)	USA / 1 year / 2003/2004	Implementing teaching materials to improve instruction	40 elementary schools, 99 second-, fourth-, and sixth-grade teachers, 1,971 students	c) .49*
5. Teacher-Led Math Inquiry (Hull et al., 2018)	Belize / 1 year / 2011/2012	Examining an effect of compound intervention on the students' mathematical skills	24 elementary and middle schools, 282 first- to eighth-grade teachers, 6,576 students	c) .27
6. Classroom Connectivity in Mathematics and Science Achievement (Pape et al., 2012, Irving et al., 2016)	USA / 1 year / 2005/2006	Using classroom connectivity technology for formative assessment	n/a schools, 82 ninth-grade teachers, 1,224 students	c) .27
7. Lesson Study (Lewis and Perry 2017)	USA / 12 weeks / 2009/2010	Investigating effectiveness of lessons supported by resource kits	39 elementary schools, 213 second- to fifth-grade teachers, 1,162 students	c) .49*

Title of program (Reference)	Country / Study duration / School year(s)	Aim	Baseline sample size	Type of measure, effect size†
8. Enhancing Missouri's Instructional Networked Teaching Strategies (Meyers et al., 2016)	USA / 3 years / 2011/2012, 2012/2013, 2013/2014	Using technology for developing student-centered instruction	60 middle schools, 100 seventh- and eighth-grade teachers, 3,072 students	a) .15*
9. Alabama Math, Science, and Technology Initiative (Newman et al., 2012)	USA / 2 years / 2006/2007, 2007/2008	Improving students' achievement by using materials, technology, and in-school support	82 elementary schools, 482 fourth- to eighth-grade teachers, 22,557 students	a) .05
10. SimCalc (Roschelle et al. 2010)	USA / 2 years / 2005/2006, 2006/2007	Using technology for learning advanced mathematics	129 middle schools, 228 seventh- and eighth-grade teachers, 2,446 students	c) .61*
11. ASSISTments (Roschelle et al., 2016)	USA / 2 years / 2012/2013, 2013/2014	Providing quality feedback and guidance for students using an online application to do homework	43 middle schools, n/a seventh-grade teachers, 2,850 students	a) .18*
12. JUMP Math (Solomon et al., 2019)	Canada / 2 years / 2013/2014, 2014/2015	Promoting a deep conceptual understanding via collaborative solving of real-world mathematical problems	41 schools, 49 fifth-grade teachers, 592 students	a) .22*
† According to Lipsey et al. (2012), educational programs with practical significance are those with effect sizes equal to or greater than: (a) .08 for broadly focused standardized tests; (b) .24 for narrowly focused standardized tests; and (c) .39 for specialized tests developed for a particular intervention. In this table we have marked statistically significant effect sizes with an asterisk (*).				

Table 2: List of papers included in the analysis.

Learning was based on prior knowledge, critical thinking, and stimulating 'creative solutions to non-routine problems and use of a variety of representations' (Heller et al., 2007, p. 2). Students had opportunities to elaborate upon their thinking, as well as discuss mathematical ideas and test their validity with peers (Heller et al., 2007; Chen et al., 2020), plus an inquiry-based approach and collaborative learning were implemented in the classroom (Meyers et al., 2016; Hull et al., 2018). Lessons utilized a fine-grained guided discovery approach, tailored according to the individual needs of students in the class (Solomon et al., 2019). Utilization of technology in the classroom also facilitated timely, supportive, and specific feedback (Pape et al., 2012; Roschelle et al., 2016), making instructional decisions easier for teachers. Teachers also used technology as an aid to clarify mathematical concepts, either by providing quality explanations using videos (Beg et al., 2019), or by supporting visualization and interaction with concrete embodiments (Roschelle et al., 2010).

Based on the papers analysed, we can conclude that changes in teaching do not always need to be comprehensive in order to lead to more effective student learning. This is confirmed in particular by the studies which reviewed the use of computer technology (Roschelle et al., 2010; Roschelle et al., 2016). These two studies found that computer-mediated mathematical content could be effective in the existing practices of most teachers, and the authors also hypothesized that some pedagogies might improve students' learning with the use of computer-mediated materials. This assumption was confirmed by Li and Ma (2010) in their meta-analysis, where they found that the use of technology has a greater effect size (i.e., 1.00 SD) when used in constructivism-based teaching.

In the selected studies, professional development hours were analyzed (Table 3). Interventions ranged from a dozen hours over a few meetings (Heller et al., 2007; Beg et al., 2019; Solomon et al., 2019; Chen et al., 2020) to 46 face-to-face sessions (a total of 240 hours) over two school years. This variation in duration suggests that effective interventions can be achieved with shorter or more intensive professional development programs and does not verify the idea that professional development must have a lengthy duration in order to be effective (Garet et al., 2001; Desimone, 2009; Darling-Hammond et al., 2017). Lynch et al. (2019) found no evidence of a positive association between professional development duration and program outcomes in their meta-analysis. If teachers and students receive additional learning incentives, the duration of professional development appears to be irrelevant (Lauer et al., 2014). The results of this analysis confirm the hypothesis posed by McEwan in an effort to answer the question: "Why are some categories [of PD intervention] apparently less effective after controlling for moderators?" (2015, p. 24). McEwan presumed that a treatment component would be more effective when combined with another complementary treatment component than either component in isolation. This assumption was proved by Snilstveit et al.'s (2016) assessment of 216 different programs in low- and middle-income countries: they found that structured pedagogy programs are the most effective of PD interventions. An ideal structured pedagogy intervention includes the following activities: "(1) teacher training, (2) ongoing teacher support, supervision and feedback, (3) provision of teacher-oriented resources or materials, [and] (4) provision of classroom learning materials" (Snilstveit et al., 2016, p. 175).

PD programme	PD duration	Structured pedagogy intervention *	Initial PD	In-year follow-up activities			Using videos	Type of teacher know.+
				Work-shops	Coach.	Online learn.		
1. eLearn	2 days	b, c	+					PCK
2. Video-based teacher PD	12 hours	a	+	+		+	+	PCK
3. ECED	5 days	a	+	+	+	+		PCK
4. MPP	8 hours	**a, b, c**	+	+			+	PCK
5. TLMI	More than 50 hours	**a, b, c**		+	+	+		CK, PCK
6. CCT	2 weeks	**a, b, c**	+	+		+	+	PCK
7. Lesson Study	7 to 42 hours	a, b		+			+	CK, PCK
8. eMINTS	240 hours	**a, b, c**	+	+	+	+	+	PCK
9. AMSTI	50 hours	**a, b, c**	+	+	+			CK, PCK
10. SimCalc	5 days	b, c	+			+	+	CK, PCK
11. ASSIST-ments	5 days	**a, b, c**	+	+	+	+		PCK
12. JUMP Math	12 hours	**a, b, c**	+	+				PCK

(See Table 2 for full versions of abbreviated titles)
* Elements of structured pedagogy intervention utilised along with teacher training: a) on-site teacher support, supervision, and feedback (multiple teacher observation and giving feedback to the teacher about his classroom action); b) resources for teachers (lesson plans, activity guides and materials, making teaching aids, etc.); and c) classroom learning materials (flash-cards, wallcharts, textbooks, workbooks, storybooks or technology etc.) (Snilstveit et al. 2016). In studies that possess all elements of structured pedagogy intervention, letters a, b, and c are marked in bold.
+ Type of teacher knowledge: Content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical knowledge (PK).

Table 3: Representation of professional development (PD) elements in effective programs

Indeed, this finding was precisely the case for seven of the 12 programs listed in Table 2 which, in addition to teacher training courses and onsite support, include teacher resources, classroom materials, and technology (e.g., paper curricula, quizzes, teaching guides for lessons, videos for teachers and students, mathematics tasks for teachers and students, and visualizations and interactive technologies). Video-based teacher professional development (Chen et al., 2020) and the ECED program (Early et al., 2016) did not include materials for teachers and students, but changes in teaching were achieved through well-designed and guided PD. In Lesson Study, the teachers cooperated without guidance from coaches, mentors, or project researchers, however they did help each other to follow a detailed resource kit. The eLearn (Beg et al., 2019) and SimCalc (Roschelle et al., 2010) projects did not provide ongoing teacher support: the initial PD they received was sufficient for learning how to use the relevant technologies. While this technology-based "streamlining' may facilitate short-term achievements, more comprehensive and lasting changes require more intensive PD in order 'to sustain and expand implementations across many years" (Roschelle et al., 2010, p. 872).

Most studies (seven out of 12) reported the use of online learning and resources in teacher PD, although merely as an addition to face-to-face PD. Thus, none of the selected programs were fully based on online PD. We assume that a key reason for the dominant role of face-to-face professional learning is related to the importance of pedagogical content knowledge (PCK), on which all of the interventions were focused. PCK is rooted in adult learning principles: intrinsic motivation; self-direction; metacognition; solving practical problems idiosyncratically related to the learner; participating in communities of practitioners; deepening understandings of professional contexts; disclosing oppressive structures and practices; and transforming habits of the mind by becoming critically reflective (Lave & Wenger, 1991; Chan, 2010; Knowles, 2015). Although there are applications which can facilitate such learning in a digital space (e.g. Moodle, Zoom, and Microsoft Teams), it seems that in-person communication remains crucial in the professional learning of mathematics teachers — which (in all studies) was achieved through summer institutes or follow-up PD meetings.

Coaching was used to supplement initial teacher education in five interventions, especially those that required significant changes in teaching (e.g. ECED Math and Literacy Matters, Teacher-Led Math Inquiry, eMINTS, and AMSTI). The coach involved in this role was usually a teacher from the school where the intervention had been conducted, and who had received additional training to serve as a PD team liaison. The coaching component was accomplished both in-person and remotely. Despite the benefits that coaching provides to teachers (Campbell & Malkus, 2011; Cordingley & Buckler 2012; Darling-Hammond et al., 2017), it is a method used **in under half of the 12 programs**. This lack of utilisation may mean that the expected changes can be achieved without this component, if other aspects of the intervention allow teachers to introduce the expected changes in their teaching. However, coaching does appear to have been an important support in implementing planned changes in teaching: the results of Kraft et al. s' (2018) meta-analysis found that coaching has a significant effect on teachers' instructional practice (i.e., 0.49 SD) and on students' academic performance (i.e., 0.18 SD).

We identified video usage in six of the programmes — videos were used to communicate project, teaching, and student activities, as well as to gather qualitative data on teaching and PD in experimental studies. In the only video-based teacher PD program (Chen et al., 2020), the emphasis was on training teachers to use academically productive talk (APT) in their mathematics teaching. The Classroom Discourse Analyzer (CDA) application enabled teachers to visualize class discussions in three ways: multiple representation (e.g., teaching videos, transcripts, and visualization of the APT moves); interactive visualization (e.g., frequency of APT moves); and contextualized evidence (e.g., observation of APT moves in a certain segment of teaching). Using videos and the CDA application allows teachers to focus on their teaching which contributes to better reflection, with "CDA [being a] tool [which] can help teachers recognize how their teaching resembles or differs from one another, which empowers evidence-based discussions and collaborative learning" (Chen et al., 2020, p. 29). All three ways of watching videos of teaching — "viewing videos of unknown teacher activity", "viewing videos of peer activity", and "viewing videos of one's own professional practice" (Gaudin & Chaliès, 2015) — were only used in the Lesson Study program. Participants in this program had the opportunity to learn by watching and discussing videos of experienced Japanese teachers. In the planning phase, teachers were required to prepare and deliver a research lesson. These research lessons were recorded, reflected on in lesson study teams, and periodically mailed as video data cards to researchers. However, none of the interventions used videos of mathematics teaching practices in an online context, which may prompt us to further explore this approach.

All of the PD programmes focused on teachers' pedagogical content knowledge, plus four studies detailed efforts to improve their content knowledge (CK). The fact that in the studies analysed a greater emphasis was placed on PCK than on CK is consistent with the conclusion reached by Baumert et al. (2010), who determined that teachers' pedagogical content knowledge better predicts the mathematics outcomes of ninth-grade students than their content knowledge does. It should be noted here that none of the programs focused on general pedagogical knowledge (PK), suggesting that focussing on teachers' PCK and CK may be sufficient for increasing students' mathematics outcomes. Nevertheless, it would be wrong to conclude that general pedagogical knowledge is not important for mathematics teachers; conversely, PK has proven to be essential in students' assessment of teaching quality in vocational schools in Austria (König & Pflanzl, 2016) and a positive predictor of learning support in Germany (Baier et al., 2019). In addition, general pedagogical knowledge is an important prerequisite for the improved professionalization of the teaching vocation (Guerriero, 2017) — therefore, it is important to explore how PK may be incorporated into the professional development of mathematics teachers.

A carefully considered connection of all the elements listed in Table 2 in the intensive three-year eMINTS professional development program (Meyers et al., 2016) has led to significant improvements in students' mathematics outcomes. In this program, PD specialists provided teachers with coaching, communities of practice, and online courses. Furthermore, another important aspect of this program was school leadership supporting "eMINTS implementation and maintain[ing] a schoolwide learning environment for teachers" (Meyers et al., 2016, p. 5). Instructional changes were focused on: collaborative and inquiry-based learning; strategies that best meet learners' needs and help them learn through reflection and metacognition; using multiple data sources to present mathematics content; feedback from assessments; and technology integration. The intervention enabled more quality learning for both teachers and students, appearing to be a winning combination.

To enhance mathematics learning outcomes for students, it appears that professional development interventions should provide on-site teacher support, mentoring, and feedback. Moreover, providing teachers with some form of teacher-focused resources and classroom learning materials would be of great assistance to those attempting to implement new instructional practices. In any case, it seems unrealistic to expect an intervention to be effective if it relies solely on a one-time PD intervention conducted in the summer preceding its first term of implementation. The utilization of technology has been shown to be essential for student learning. Two programs yielded significant positive results (Roschelle et al., 2010; Roschelle et al., 2016), in which educational software played a central role in student learning, despite relatively minor changes in teaching practices. Although these interventions did not necessitate substantial changes in instruction, students' mathematics performance improved due to the success of computer applications in facilitating deeper learning. The findings of these studies may suggest that changes in instruction can be implemented incrementally, beginning with less complex interventions and progressing to those that are more complex and demanding. When teachers see the benefits of new approaches based on the results of their own practice, it is feasible to continue with a more intensive form of professional development that will prepare them for a deeper understanding and the development of innovative practices. This recommendation is consistent with Guskey's conclusion that "significant change in teachers' attitudes and beliefs occurs predominantly after they gain evidence of improvements in student learning" (2002, p. 383). Although the majority of PD was conducted face-to-face with leaders and other program participants, online platforms and videos were used as a supplement to teacher training rather than as a central component, indicating that effective, modern PD does not necessarily require the use of technology. A recommendation from our review would be to investigate whether and how technology in professional development can meaningfully assist mathematics teachers in improving student learning outcomes.

CONCLUSIONS

Professional development can be structured in various ways to enhance its effectiveness. The provision of high-quality initial PD and subsequent follow-up workshops can be effectively enhanced through the use of structured pedagogical intervention, coaching, video resources, and online learning platforms. The primary objective, as emphasized throughout this review, is to assist teachers in implementing instructional changes that improve student learning outcomes. Using the elements that have been shown to be effective in PD interventions, we propose a strategic path for teachers who are committed to enhancing their pedagogical practices despite a potential lack of high-quality opportunities for professional development: Together with school leadership, a teacher should create a community of learning that concentrates on improving his/her teaching practice. The teacher should expose students in his or her lessons to an active teaching-learning process. The teacher should use a fine-grained, individualized guided discovery approach, provide optimal challenges for each student, and could use technology for timely, supportive, and specific feedback. In addition, the teacher should videotape his/ her lessons in order to evaluate the instructional strategies employed and determine whether or not these strategies promoted student engagement and active participation. This analysis of the video should be conducted in conjunction with the school's established community of learning. Protocols from effective PDs should be utilized in this process, as they can provide precise feedback on instructional practices and student engagement. In addition, the teacher should establish a partnership with an expert teacher who could serve as a coach. If possible, from his or her own school; if not, from a nearby school. Lastly, teachers must consistently pursue the improvement of PCK, but school leadership plays a crucial role in providing support for such efforts.

STUDY LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

The studies in this systematic review mainly relate to the USA because eight of the 12 studies analysed were conducted there. Recognizing that teaching is culturally diverse and that the United States lacks a national curriculum (Kennedy, 2016), we have included international research. While analysing studies from various nations, we discovered a number of significant sociocultural distinctions, but they are not so substantial that research from one nation cannot inform practice and policy in others. The fact that we focused solely on randomized experiments for the purposes of our systematic review may be considered an advantage, but it is also a limitation. Quantitative results are analysed and presented in experimental studies, whereas qualitative data on intervention implementation is rarely used. However, qualitative data are frequently required to fully comprehend and replicate the effectiveness of an intervention across educational contexts. To learn more about the actual implementation and application of the programs in our analysis, we consulted other published papers, particularly qualitative research, to gather more information about the conducted PD activities (e.g., Lewis & Perry, 2014) or teaching (Bell & Pape, 2012). This additional step in our literature review demonstrates that researchers must conduct and publish qualitative and experimental research in addition to quantitative research or employ a mixed-method research design.

Acknowledgments

This work has been fully supported by the Croatian Science Foundation under the project IP-2018-01-8363.

References[1]

[1] Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education, 27*(1), 10–20. https://doi.org/10.1016/j.tate.2010.08.007

[2] Baier, F., Decker, A.-T., Voss, T., Kleickmann, T., Klusmann, U., & Kunter, M. (2019). What makes a good teacher? The relative importance of mathematics teachers' cognitive ability, personality, knowledge, beliefs, and motivation for instructional quality. *British Journal of Educational Psychology*, *89*, 767-786. https://doi.org/10.1111/bjep.12256

[3] Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics, 102*(1), 1–8. https://doi.org/10.1007/s10649-019-09908-4

[4] Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

[5] Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180, https://doi.org/10.3102/0002831209345157

[6] Bautista, A., & Oretga-Ruiz, R. (2017). Teacher professional development: International perspectives and approaches. *Psychology, Society, & Education*, *7*(3). 240–251, https://doi.org/10.25115/psye.v7i3.1020

[7] *Beg, S. A., Lucas, A. M., Halim, W., & Saif, U. (2019). *Beyond the basics: Improving post-primary content delivery through classroom technology*. NBER Working Paper No. 25704. National Bureau of Economic Research. https://doi.org/10.3386/w25704

[8] Bell, C. V., & Pape, S. J. (2012). Scaffolding students' opportunities to learn mathematics through social interactions. *Mathematics Education Research Journal*, *24*(4), 423–445. https://doi.org/10.1007/s13394-012-0048-1

[9] Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal, 111*(3), 430–454, https://doi.org/10.1086/657654

[10] Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., … Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, *45*(4), 419–459. https://doi.org/10.5951/jresematheduc.45.4.0419

[11] Chan, S. (2010). Applications of andragogy in multi-disciplined teaching and learning. *Journal of Adult Education*, 39(2), 25–35,

[12] *Chen, G., Chan, C. K. K., Chan, K. K. H., Clarke, S. N., & Resnick, L. B. (2020). Efficacy of video-based teacher professional development for increasing classroom discourse and

[1] References marked with an asterisk indicate studies included in the review.

student learning. *Journal of the Learning Sciences, 29*(4–5), 642–680. https://doi.org/10.1080/10508406.2020.1783269

[13] Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

[14] Cordingley, P., & Buckler, N. (2012). Mentoring and coaching for teachers' continuing professional development. In S. J. Fletcher & C. A. Mullen (Eds.), *The SAGE handbook of mentoring and coaching in education* (pp. 215–227). SAGE. https://doi.org/10.4135/9781446247549

[15] Day, C., & Leitch, R. (2007). The Continuing Professional Development of Teachers: Issues of Coherence, Cohesion and Effectiveness. In T. Townsend (Ed.), *International Handbook of School Effectiveness and Improvement* (pp. 707–726). Springer. https://doi.org/10.1007/978-1-4020-5747-2_38

[16] Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/Effective_Teacher_Professional_Development_REPORT.pdf

[17] Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

[18] *Early, D. M., Berg, J. K., Alicea, S., Si, Y., Aber, J. L., Ryan, R. M., & Deci, E. L. (2016). The impact of Every Classroom, Every Day on high school student achievement: Results from a school-randomized trial. *Journal of Research on Educational Effectiveness*, *9*(1), 3–29. https://doi.org/10.1080/19345747.2015.1055638

[19] Evans, L. (2014). Leadership for professional development and learning: Enhancing our understanding of how teachers develop. *Cambridge Journal of Education, 44*(2), 179–198. https://doi.org/10.1080/0305764X.2013.860083

[20] Fraser, C., Kennedy, A., Reid, L., & Mckinney, S. (2007). Teachers' continuing professional development: Contested concepts, understandings and models. *Journal of In-Service Education, 33*(2), 153–169. https://doi.org/10.1080/13674580701292913

[21] Fullan, M., & Hargreaves, A. (2016). *Bringing the profession back in: Call to action*. Learning Forward. Learning Forward. https://learningforward.org/wp-content/uploads/2017/08/bringing-the-profession-back-in.pdf

[22] Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945. https://doi.org/10.3102/00028312038004915

[23] Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review, 16*, 41–67. https://doi.org/10.1016/j.edurev.2015.06.001

[24] Gough, D., Oliver, S., & Thomas, J. (Eds.). (2017). *An introduction to systematic reviews*. SAGE.

[25] Guerriero, S. (Ed.). (2017). *Pedagogical knowledge and the changing nature of the teaching profession*. OECD Publishing. https://doi.org/10.1787/20769679

[26] Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, 8(3), (pages 381–391) https://doi.org/10.1080/135406002100000512

[27] *Heller, J. I., Curtis, D. A., Rabe-Hesketh, S., & Verboncoeur, C. J. (2007). *The effects of "Math Pathways and Pitfalls" on students' mathematics achievement: National science foundation final report*. https://eric.ed.gov/?id=ED498258

[28] Hendriks, M. A., Luyten, H., Scheerens, J., Sleegers, P., & Steen, R. (2010). *Teachers' professional development: Europe in international comparison*. Office for Official Publications of the European Union. https://doi.org/10.2766/63494

[29] Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

[30] *Hull, D. M., Hinerman, K. M., Ferguson, S. L., Chen, Q., & Näslund-Hadley, E. I. (2018). Teacher-led math inquiry: A cluster randomized trial in Belize. *Educational Evaluation and Policy Analysis*, *40*(3), 336–358. https://doi.org/10.3102/0162373718768430

[31] *Irving, K. E., Pape, S. J., Owens, D. T., Abrahamson, L., Silver, D., & Sanalan, V. A. (2016). Classroom connectivity and Algebra 1 achievement: A three-year longitudinal study. *Journal of Computers in Mathematics and Science Teaching*, *35*(2), 131–151.

[32] Kennedy, M. M. (1999). Form and substance in mathematics and science professional development. *NISE Brief*, *3*(2), 1–8. https://files.eric.ed.gov/fulltext/ED435552.pdf

[33] Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

[34] Knowles, M. S., Holton III, E. F., & Swanson, R. A. (2015). *The adult learner: The definitive classic in adult education and human resource development*. Routledge.

[35] König, J., & Pflanzl, B. (2016). Is teacher knowledge associated with performance? On the relationship between teachers' general pedagogical knowledge and instructional quality. *European Journal of Teacher Education*, *39*(4), 419–436. https://doi.org/10.1080/02619768.2016.1214128

[36] Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

[37] Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588. https://doi.org/10.3102/0034654318759268

[38] Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM*, *40*(5), 873–892. https://doi.org/10.1007/s11858-008-0141-9

[39] Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805–820. https://doi.org/10.1037/a0032583

[40] Lauer, P. A., Christopher, D. E., Firpo-Triplett, R., & Buchting, F. (2014). The impact of short-term professional development on participant outcomes: A review of the literature.

Professional Development in Education, *40*(2), 207–227. https://doi.org/10.1080/19415257.2013.776619

[41] Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

[42] Lewis, C., & Perry, R. (2014). Lesson study with mathematical resources: A sustainable model for locally-led teacher professional learning. *Mathematics Teacher Education and Development*, *16*(1), 22-42.

[43] *Lewis, C., & Perry, R. (2017). Lesson study to scale up research-based knowledge: A randomized, controlled trial of fractions learning. *Journal for Research in Mathematics Education*, *48*(3), 261–299. https://doi.org/10.5951/jresematheduc.48.3.0261

[44] Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, *22*(3), 215–243. https://doi.org/10.1007/s10648-010-9125-8

[45] Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., … Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf

[46] Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, *41*(3), 260–293. https://doi.org/10.3102/0162373719849044

[47] Marcelo, C. (2009). Professional development of teachers: Past and future. *Sísifo: Educational Sciences Journal*, *8*, 5–19.

[48] McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, *85*(3), 353–394. https://doi.org/10.3102/0034654314553127

[49] *Meyers, C. V., Molefe, A., Brandt, W. C., Zhu, B., & Dhillon, S. (2016). Impact results of the eMINTS professional development validation study. *Educational Evaluation and Policy Analysis*, *38*(3), 455–476. https://doi.org/10.3102/0162373716638446

[50] *Newman, D., Finney, P. B., Bell, S., Turner, H., Jaciw, A. P., Zacamy, J. L., & Gould, L. F. (2012). *Evaluation of the effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI): Final report* (NCEE 2012-4008). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_20124008.pdf

[51] Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 1-10. https://doi.org/10.1186/s13643-016-0384-4

[52] *Pape, S. J., Irving, K. E., Owens, D. T., Boscardin, C. K., Sanalan, V. A., Abrahamson, A. L., … Silver, D. (2012). Classroom connectivity in Algebra I classrooms: Results of a randomized control trial. *Effective Education*, *4*(2), 169–189. https://doi.org/10.1080/19415532.2013.841059

[53] Pellegrini, M., Lake, C., Inns, A., & Slavin, R. E. (2018). *Effective programs in elementary mathematics: A best-evidence synthesis* [Best Evidence Encyclopedia]. http://www.bestevidence.org/word/elem_math_Oct_8_2018.pdf

[54] Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Wiley-Blackwell.

[55] Polanin, J. R., Maynard, B. R., & Dell, N. A. (2017). Overviews in education research: A systematic review and analysis. *Review of Educational Research*, *87*(1), 172–203. https://doi.org/10.3102/0034654316631117

[56] *Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open*, *2*(4). https://doi.org/10.1177/2332858416673968

[57] *Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., … Gallagher, L. P. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal*, *47*(4), 833–878. https://doi.org/10.3102/0002831210367426

[58] Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

[59] Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A Best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70*(1), 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

[60] Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, *37*(1), 5–14. https://doi.org/10.3102/0013189X08314117

[61] Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, *31*(4), 500–506. https://doi.org/10.3102/0162373709352369

[62] Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., … Jimenez, E. (2016). *The impact of education programmes on learning and school participation in low- and middle-income countries* (3ie Systematic Review 24). International Initiative for Impact Evaluation (3ie). https://doi.org/10.23846/SRS007

[63] *Solomon, T., Dupuis, A., O'Hara, A., Hockenberry, M.-N., Lam, J., Goco, G., Ferguson, B., & Tannock, R. (2019). A cluster-randomized controlled trial of the effectiveness of the JUMP Math program of math instruction for improving elementary math achievement. *PLOS ONE, 14*(10), 1-36. https://doi.org/10.1371/journal.pone.022304

[64] What Works Clearinghouse™. (2020). *Standards Handbook* (Version 4.1). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf