Progress in Science Education
2023, Vol. 6, No. 2, 47-62
ISSN 2405-6057
DOI.10.25321/prise.2023.1401

Special Section
Proceedings of the Swiss Doctoral School of Science Education 2022

Registered Research Plan

A Professional Development Program to Foster Science Teachers' Professional Competence, Enhance Classroom Practice, and Improve Student Outcomes Related to Scientific Reasoning

Richard Sannert[1], Moritz Krell[1]

Received: October 2022 / Accepted: May 2023

Structured Abstract

Background: Developing scientific reasoning competencies that enable students to understand how scientific knowledge develops is considered an important goal of science education. Thus, the literature recommends engaging students in scientific experimentation and reflecting explicitly on the procedural and epistemic knowledge involved. However, science teachers in middle schools often do not provide enough opportunities for their students to improve their scientific reasoning competencies. This lack of learning opportunities seems to be a consequence of teachers' professional competence in this area not yet being sufficiently developed. Therefore, the call for effective and continuous professional development (PD) programs has been made.
Purpose: The planned PD program is intended to (1) foster science teachers' professional competence, (2) enhance classroom practice, and (3) improve student outcomes related to scientific reasoning. The study firstly aims to measure all three kinds of effects of the PD program. Secondly, the study aims to identify learning pathways and learning obstacles in the teachers' implementation of explicit instruction in experimentation.
Sample/setting: In total, 20 in-service biology teachers from the federal state of Schleswig-Holstein in Germany who are teaching in classes from Grade 5 to Grade 10 will be recruited to take part in the PD program. The PD program adopts content-related design features (e.g., developed lessons, modeled inquiry) that, from our point of view, can be considered as specifications of the more generic design of effective PD programs (e.g., content focus, active learning).
Design and Methods: A pre-post-follow-up study is planned to evaluate the effect of the PD program on teachers' content knowledge (CK) and pedagogical content knowledge (PCK), as well as on teachers' beliefs and their self-efficacy towards teaching experimentation. Furthermore, we aim to capture changes in classroom practice via direct observation. For the student outcomes, we will take scientific reasoning competencies and motivation into account and compare the development with a comparison group. Learning pathways and obstacles will be identified through the qualitative analysis of observation protocols, transcripts from the videotaped lessons, lesson plans, and other qualitative data derived from the PD program.
Conclusions: Few studies to date have examined the effect of a PD program on teachers' professional competence, classroom practice, and student outcomes. The present study aims to contribute to closing this research gap. Furthermore, knowledge about teachers' learning pathways and obstacles to implementation will provide insights into how to better support science teachers with a PD program that aims to improve students' scientific reasoning competencies through explicit instruction in experimentation.
Keywords: *professional development program, scientific reasoning, experimentation, in-service teachers*

[1]IPN - Leibniz Institute for Science and Mathematics Education
✉ sannert@leibniz-ipn.de

1 Introduction

Most of the challenges today's societies face (e.g., climate change, anti-scientific information, or—recently—the COVID-19 pandemic), as well as many issues individuals face in their everyday lives (e.g., vaccination, nutrition, electric mobility), are closely related to science (Hameleers & Van der Meer, 2021; Sharon & Baram-Tsabari, 2020). Developing scientific reasoning competencies that enable students to understand how scientific knowledge develops and to critically evaluate scientific findings is therefore considered an important goal of science education (KMK, 2020; NRC, 2012; OECD; 2017). To achieve this goal, science teaching should not only emphasize content knowledge but also, explicitly, procedural knowledge (knowing *how*) and epistemic knowledge (knowing *why*) (Krell et al., 2022; Osborne, 2014; Vorholzer et al., 2020). One way to address scientific reasoning competencies in science instruction is through experimentation (Kind & Osborne, 2016). However, science teachers in middle schools often do not provide enough opportunities for their students to improve their scientific reasoning competencies through experimentation (Büssing et al., 2022; Capps & Crawford, 2013). This is critical because students are not gaining the competencies necessary to meet current and future challenges.

It is reasonable to assume that science teachers' professional competence plays a key role in providing students with appropriate learning opportunities (Capps & Crawford, 2013; Lederman & Lederman, 2019). One component of teachers' professional competence is their professional knowledge. Teachers' professional knowledge can positively influence the quality of classroom practice (Kelcey et al., 2019; Kind et al., 2022). In terms of scientific reasoning, it has been shown that science teachers who have comprehensive professional knowledge about experimentation are more likely to provide appropriate learning opportunities for their students (Capps et al., 2012). Another component of teachers' professional competence is their beliefs (Baumert & Kunter, 2013a). Teachers' beliefs could explain a lack of classroom practice that promotes experimentation and makes procedural and epistemic knowledge explicit (Enzingmüller & Prechtl, 2021; Petermann & Vorholzer, 2022). Given the need to foster science teachers' professional competence to enable them to enhance their classroom practice and to improve students' outcomes related to scientific reasoning, the call for effective and continuous professional development (PD) has been made (Capps et al., 2016; Osborne et al., 2022).

As a result of research in the field of teachers' PD, researchers have reached an evidence-based consensus on a set of generic design features that PD programs should have in order to be effective (e.g., Darling-Hammond et al., 2017; Desimone, 2009; Lipowsky & Rzejak, 2019). Furthermore, the literature describes content-related design features of effective PD programs for inquiry-based teaching (Capps et al., 2012; Ramnarain et al., 2022). For this study, we used these content-related design features to design a PD program that is intended to foster science teachers' professional competence, enhance classroom practice, and improve student outcomes related to scientific reasoning.

In recent years, many studies have examined the effect of PD programs on either teachers' professional competence, classroom practice, or student outcomes. However, studies that have considered all three kinds of effects are rare (Gess-Newsome et al., 2019; Lipowsky & Rzejak, 2019). This room for improvement applies in particular to PD programs related to experimentation (Ramnarain et al., 2022). With our study, we aim to contribute to closing this gap in the research by measuring all three kinds of effects of the planned PD program. Few studies on PD programs to date have focused on the learning processes of the teachers participating in a PD program (Goldsmith, 2014; Stahnke et al., 2016). This is critical because knowledge about these teacher-level learning processes (e.g., typical starting points, learning pathways, or learning obstacles when integrating PD content into classroom practice) can lead to more detailed knowledge about how content-related PD programs should be designed to be effective (Prediger et al., 2017; Roesken-Winter et al., 2021). Therefore, in our study, we plan to investigate the learning processes of the participating teachers. This paper presents the planned empirical investigation of the PD program's effectiveness and of the teachers' learning processes.

2 Research Background

We first summarize the generic design features of effective PD programs (Section 2.1) and show their connection with the content-related design features of effective PD programs for inquiry-based teaching (Section 2.2). These form the basis for the PD program developed in this study, which is intended to improve students' scientific reasoning competencies through experimentation. After presenting a framework for the evaluation of our PD program (Section 2.3), we identify room for improvement in studies on effective PD programs (Section 2.4) and present our research questions (Section 2.5).

2.1 Generic Design Features of Effective PD Programs

In the last decades, researchers in the field of teachers' PD have reached an evidence-based consensus on a set of generic design features that PD programs should have in order to be effective (Desimone, 2009):

(1) An effective PD program has a clear *content focus* (Desimone, 2009), providing possibilities for the teachers to increase their content knowledge (CK), which then allows them to teach a topic more effectively (Baumert & Kunter, 2013b; Kind et al., 2022). Knowledge about how students can learn this content, which refers to pedagogical content

knowledge (PCK; Park & Oliver, 2008; Shulman, 1986), is also considered part of this content focus by some authors (e.g., Darling-Hammond et al., 2017).
(2) The participating teachers should be encouraged to engage in *active learning*, including, for example, classroom observations, analyzing videotaped classroom lessons, or discussions (Desimone, 2009; Darling-Hammond et al., 2017). This promotes the perception among teachers that learning is an active process (Borko et al., 2010) and it can reveal how well the teachers understand the PD content (Penuel et al., 2007).
(3) *Coherence* between a PD program and the relevant national educational standards and curricula as well as the teachers' knowledge and beliefs increases the effectiveness of the PD program (Desimone, 2009). However, teachers' knowledge and beliefs should develop in the intended direction through the teachers' participation in a PD program (Desimone, 2009). Furthermore, ensuring that the PD program is aligned with the teachers' classroom practice makes an integration of PD content more likely (Desimone & Garet, 2015).
(4) An adequate *duration*, including both a minimum total time of the PD program and a certain span of time over which the units of the PD program are spread, is considered necessary to achieve a change in teachers' professional competence, their classroom practice, and their students' outcomes (Desimone, 2009). The minimum total time recommended in the literature ranges from 14 hr (Yoon et al., 2007) to 20 hr (Desimone, 2009). Nevertheless, there is a consensus that one-shot workshops, although often still conducted, do not show lasting effects (Lipowsky & Rezjak, 2019). However, both the interaction between the duration and other design features of the PD program and the concrete operationalization of those design features seem to be more crucial factors than the duration per se (Desimone & Garet, 2015; Lipowsky & Rzejak, 2019).
(5) Finally, the *collective participation* of teachers from the same school and/or the same grade is seen as a design feature of effective PD programs because it can promote collegial interaction and the motivation to solve problems together (Desimone, 2009; Yang et al., 2020).

2.2 Content-Related Design Features of Effective PD Programs for Inquiry-Based Teaching

In addition to these rather generic design features, more content-related design features are particularly important in designing effective PD programs for a specific content area (Prediger et al., 2017; Roesken-Winter et al., 2021). In two related reviews, 11 crucial content-related design features of effective PD programs for inquiry-based teaching (e.g., teaching experimentation) were extracted from the literature (Capps et al., 2012; Ramnarain et al., 2022): *content knowledge, authentic experience, developed lessons, modeled inquiry, reflection, coherency, transference, total time, extended support, collaborative community,* and *blended learning*. Although these content-related design features have been identified in independent literature reviews, from our point of view, they can be considered as specifications of the generic design features (Tab. 1).
Content knowledge in relation to inquiry-based teaching means, for example, knowledge about the nature of science, scientific inquiry, or scientific concepts (Capps et al., 2012). Thereby, this design feature is conceptually close to the generic design feature of *content focus* (Section 2.1). However, while CK and PCK is covered by the design feature *content focus* in the generic design features of effective PD programs (Desimone, 2009; Darling-Hammond et al., 2017; Section 2.1), PCK is not explicitly addressed in the content-related design feature *content knowledge* of effective PD programs for inquiry-based teaching (Capps et al., 2012; Ramnarain et al., 2022). Given the relevance of PCK for successful teaching (Baumert & Kunter, 2013b; Kulgemeyer & Riese, 2018), we think fostering participants' PCK should be an aim of any PD program to be effective. The following four design features correspond to the second generic design feature, *active learning*, because they cannot be effective if the science teachers passively listen to a lecture (Section 2.1). Instead, they require the commitment and active participation of the science teachers during the PD program: *Authentic experience* means that the science teachers themselves design and carry out scientific investigations, sometimes together with scientists. The design feature of *developed lessons* requires science teachers to design inquiry-based lessons for their students or to adapt their own teaching material. Through *modeled inquiry*, the science teachers engage in scientific inquiry from the perspective of their students. In our opinion, *developed lessons* and *modeled inquiry* can be seen as ways to promote PCK in a PD program because they address knowledge of instructional strategies and knowledge of student understanding. Another design feature, time for *reflection*, gives science teachers possibilities to reflect about their experience, for example, by journaling alone or by discussing their experience in groups. *Coherency* refers to the alignment of the PD content with national educational standards and curricula, and *transference* concerns explicit discussions about how the curriculum is implemented in science teachers' classrooms; both design features correspond to the generic design feature of *coherence* (Section 2.1). The *total time* of a PD program is equivalent to the *duration* of the PD program (Section 2.1). *Extended support* for the participants is provided by, in contrast to a one-shot workshop, a PD program that consists of several workshops, classroom observations, or remote support throughout the year; these aspects were also included in Desimone's (2009) definition of the generic *duration* feature (Section 2.1). The potential to create a *collaborative community* during and outside the PD program, with possibilities for group activities (e.g., planning lessons together) or supporting each other when integrating the PD content after the PD program, refers to the generic design feature described as *collective participation* (Section 2.1). The last content-related design feature of an effective PD program for inquiry-based teaching is *blended learning*, which means that parts of the PD program are promoted online. There is no counterpart for this design feature in the list of generic design features.

Tab. 1. Generic and content-related design features of effective PD programs

Generic design features of a PD program (Desimone, 2009)	Content-related design features of a PD program for inquiry-based teaching (Capps et al., 2012; Ramnarain et al., 2022)
Content focus	Content knowledge
Active learning	Authentic experience Developed lessons Modeled inquiry Reflection
Coherence	Coherency Transference
Duration	Total time Extended support
Collective participation	Collaborative community
	Blended learning

2.3 Framework for the Evaluation of the PD Program

Four kinds of effects are usually considered in the evaluation of a PD program: *participants' reactions, teachers' knowledge and beliefs*, *classroom practice*, and *student outcomes*. These different kinds of effects are sometimes described as "levels" (e.g., Kirkpatrick & Kirkpatrick, 1994; Lipowsky & Rzejak, 2019). However, there does not always appear to be a direct path from the PD program to teachers' knowledge and beliefs to classroom practice to student outcomes. Instead, there are multiple ways in which a PD program can affect teachers' knowledge and beliefs, classroom practice, and student outcomes (Desimone, 2009; Goldsmith et al., 2014). For example, Guskey (2002) argued that teachers' beliefs about teaching could change over time through positive feedback as a result of changes in classroom practice, which could then be followed by positive student outcomes. In addition to that, the same PD program may also have different effects on the individual teachers' knowledge and beliefs as well as on their classroom practice (Desimone & Garet, 2015).

For this study, we focused on three kinds of effects of the PD program (Fig. 1): effects (1) on *teachers' knowledge and beliefs*, (2) on *classroom practice*, and (3) on *student outcomes*. We decided not to include *participants' reactions* in our evaluation for reasons of test economy and because it is not a valid indicator of changes in classroom practice (Lipowsky & Rzejak, 2019).

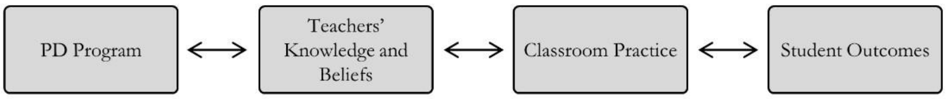

Fig. 1. Kinds of effects of the PD program intended to foster science teachers' knowledge and beliefs, enhance classroom practice, and improve student outcomes related to scientific reasoning (adapted from Desimone, 2009).

(1) In terms of teachers' professional knowledge, two knowledge domains are typically addressed in content-related PD programs: CK and PCK (e.g., Gess-Newsome et al., 2019; Yang et al., 2020). CK describes teachers' knowledge of the subject matter taught (Kleickmann et al., 2013). Several studies found that teachers' CK was a significant predictor of classroom practice (Kelcey et al., 2019; Kind et al., 2022) and student outcomes (Diamond et al., 2014; Gess-Newsome et al., 2019). Furthermore, several authors have suggested that the effect of CK on classroom practice and student outcomes is indirect and is mediated by teachers' PCK (Baumert & Kunter, 2013b; Kulgemeyer & Riese, 2018). PCK describes knowledge about teaching (Shulman, 1986) and has two central facets, *knowledge of student understanding* (KSU) and *knowledge of instructional strategies and representations* (KISR; Park & Oliver, 2008; Park & Chen, 2012). PCK has been identified as an important component of teachers' professional knowledge that impacts classroom practice (Garet et al., 2011) and student outcomes (Keller et al., 2017). In addition to these domains of teachers' professional knowledge, teachers' beliefs are assumed to be an important component of teachers' professional competence (Baumert & Kunter, 2013a; Hoy et al., 2006). Teachers' beliefs can be described as "an individual's judgement of the truth or falsity of a proposition" (Pajares, 1992, p. 316); they are considered to be somewhat changeable over time and/or through experiences but to be mostly stable within an individual (Fives & Gill, 2015). Teachers' beliefs are seen as a crucial factor for changes in teachers' classroom practice (Buehl & Becks, 2015; Lumpe et al., 2012). Teachers' beliefs can be object-related or self-related (Hoy et al., 2006). Some authors found a positive relation between science teachers' beliefs about teaching science (i.e., object-related beliefs; e.g., Enzingmüller & Prechtl, 2021; Lotter et al., 2020) and their classroom practice.

Additionally, teachers' self-efficacy beliefs have been shown to impact student learning through teachers' perceptions of their own skills, their expectations of their students, and the learning environments they establish in their classes (Bandura, 1997; Barni et al., 2019). Teachers with high self-efficacy beliefs have been found to elicit greater student outcomes (Goddard et al., 2000; Mohamadi & Asadzadeh, 2012). Against this theoretical background, we argue that the effects of PD programs on teachers' CK and PCK, as well as on their beliefs about teaching and on their self-efficacy beliefs, are relevant when evaluating PD programs related to scientific reasoning.

(2) Although teachers' knowledge and beliefs can positively affect classroom practice, these factors cannot fully explain the impact of PD programs on classroom practice (Lederman & Lederman, 2019; Yang et al., 2020). Teachers' classroom practice is seen as an important factor for student outcomes (Fischer et al., 2018). Previous studies found that teachers who provide more inquiry-based instruction (e.g., experimentation) give their students more learning opportunities and, thus, improve student outcomes in terms of scientific reasoning competencies (Yang et al., 2020). Additionally, engaging students in experimentation combined with explicit reflection on the procedural and epistemic knowledge involved has been found to be effective in promoting student outcomes (Vorholzer et al., 2020; Wagensveld et al., 2015). Several methods are described in the literature to measure the effects of PD programs on teachers' classroom practice (Desimone, 2009; Guskey, 2002). These range from teacher self-reports (e.g., surveys, interviews, participant reflections, or portfolios) to various forms of classroom observations (e.g., direct classroom observations, videos, or audio tapes), with the latter being considered to be more objective (Jacobs et al., 2014). Van Driel et al. (2012) stated that most studies on PD programs use self-reported data for the evaluation of the effects on classroom practice. This has been criticized because self-reported data can differ considerably from actual classroom practice (e.g., Capps et al., 2016; Hiebert & Grouws, 2007). Consequently, classroom observations should be considered in order to gain more valid information about the effects of PD programs on classroom practice (Jacobs et al., 2014).

(3) To measure student outcomes, studies in this area typically assess cognitive components such as student achievement (Gess-Newsome et al., 2019; Lederman & Lederman, 2019). However, as Zeichner (2005) called for a broader conceptualization of how to measure student outcomes, the concept of competence (Rychen & Salganik, 2003; Weinert, 2001) could be a fruitful theoretical framework for considering the cognitive components (e.g., procedural and epistemic knowledge) and, additionally, the motivational components of student outcomes when evaluating the effects of PD programs. Assessing the motivational components of student outcomes is especially useful in relation to experimentation because a positive effect of experimentation on students' motivation can be expected (Cairns & Areepattamannil, 2019). Accordingly, in this study, we plan to evaluate motivational components in addition to student achievement.

2.4 Room for Improvement in the Evaluation of PD Programs

In recent years, many studies have examined the effect of PD programs on either teachers' knowledge and beliefs, classroom practice, or student outcomes. However, studies that have considered all three kinds of effects are rare (Gess-Newsome et al., 2019; Lipowsky & Rzejak, 2019). This seems to be the case for PD programs in general and even more for content-related PD programs related to experimentation (Ramnarain et al., 2022). Therefore, there is a gap in the research in this field. With our study, we aim to contribute to closing this gap and shed light on the effects of PD programs on teachers' knowledge and beliefs, classroom practice, and student outcomes. Additionally, the design and methods of some previous studies that examined the effects of PD programs have been criticized. For example, some reports lacked basic information about the study design and/or the PD program and, thus, cannot be replicated (Goldsmith et al., 2014; Sztajn, 2011). Furthermore, in some studies, the outcome measures chosen were rather general, which resulted in incongruence between the goals of the PD program and the outcome measures (Hattie, 2009; Van Driel et al., 2012). Follow-up tests were often missing or were administered too early in the study design (Lipowsky & Rezjak, 2019), which may have led to invalid results because some effects only become apparent a certain amount of time after the PD program has been completed (e.g., one year later; Allen et al., 2011; Kennedy, 2016).

In addition to these shortcomings, little is known about teachers' learning processes and about how to help teachers to incorporate new ideas from PD programs into their actual classroom practice (Goldsmith et al., 2014; Stahnke et al., 2016). Knowledge about teachers' typical starting points, learning pathways, and obstacles they perceive in the adaptation of the PD content for their classroom practice can serve as a basis for the development of an effective content-related PD program (Prediger et al., 2016, 2017). Such knowledge could be valuable in answering the question of how to operationalize generic design features for specific content (Rösken-Winter et al., 2021), for example, by learning more about which content-related knowledge (i.e., the rather vaguely defined design feature of *content knowledge*) is important to enable teachers to address students' scientific reasoning competencies effectively in the classroom.

2.5 Research Questions

Due to the lack of robust studies measuring the effects of PD programs on teachers' knowledge and beliefs, classroom practice, and student outcomes (Gess-Newsome et al., 2019; Lipowsky & Rzejak, 2019, Ramnarain et al., 2022), as well as the lack of detailed knowledge about the learning processes of teachers who are participating in PD programs (Goldsmith et al., 2014; Stahnke et al., 2016), we aimed to design an appropriate PD program to foster science teachers' professional competence, enhance classroom practice, and improve student outcomes related to scientific reasoning and to examine its effects. Our study was guided by the following research questions:

1. What are the effects of participating in the PD program on teachers' knowledge and beliefs, the quality of classroom practice, and student outcomes related to scientific reasoning?
2. Which content-related learning pathways do teachers typically take during the PD program and which obstacles prevent implementation in classroom practice?

3 Methods

For our PD program, a well-founded PD model from mathematics teacher education, called the Problem-Solving Cycle (PSC; Koellner et al., 2007), was adapted. In the following sections, the adapted PSC is presented with the associated content-related design features for effective PD programs for inquiry-based teaching (Section 3.1). Furthermore, we describe our intended sample (Section 3.2), as well as the planned data collection (Section 3.3) and data analyses (Section 3.4). The PD program and related studies are part of the LeFEB project.

3.1 Structure of the PD Program: Adaptation of the Problem-Solving Cycle

The PSC was originally designed to assist mathematics teachers in supporting their students' mathematical reasoning competencies (Koellner et al., 2007). We decided to focus on experimentation in the context of biology as one of the scientific methods, keeping in mind that each scientific method has its own distinct procedural and epistemic knowledge (Osborne, 2014). Therefore, our PD program aimed to assist biology teachers in teaching experimentation, that is, in engaging students in experimentation and in explicitly teaching the related procedural and epistemic knowledge. The design of the PSC was influenced by both constructivist and situative perspectives (Koellner et al., 2007). Three design principles are central to the PSC: (1) teachers actively participate in the learning process, (2) teachers' own classrooms are used as a powerful context, and (3) teachers' learning is enhanced by creating a supportive professional community (Koellner et al., 2007). The PSC has already been used on a large scale in the United States and positive effects have been found on mathematics teachers' CK and PCK, classroom practice, and student outcomes (Borko et al., 2015; Jacobs et al., 2014; Koellner et al., 2011).
In the following, we present a brief description of the adapted PSC as a PD program on explicit instruction in experimentation. The original PSC, which consists of three workshops and one video recording of classroom practice, is repeated several times with the same participants and uses content that builds on previous content (Koellner et al., 2007). In contrast, our PD program consists of three workshops (3 hr each) and one video-recorded classroom lesson (1 h) after Workshop 1, that is, it has a linear structure instead of several cycles (Fig. 2). Additionally, we plan to give the participants a briefing before the program (2 hr) and a final debriefing (2 hr) after the program has been completed (Fig. 2). The duration of our PD program is 14 hr (*total time*). This is the suggested minimum duration for effective PD programs (Yoon et al., 2007). However, as discussed above, the concrete activities in the PD program might be more crucial for its effectiveness than the duration of the program (Desimone & Garet, 2015; Lipowsky & Rzejak, 2019). The whole PD program is spread over a period of 24 weeks (*extended support*).

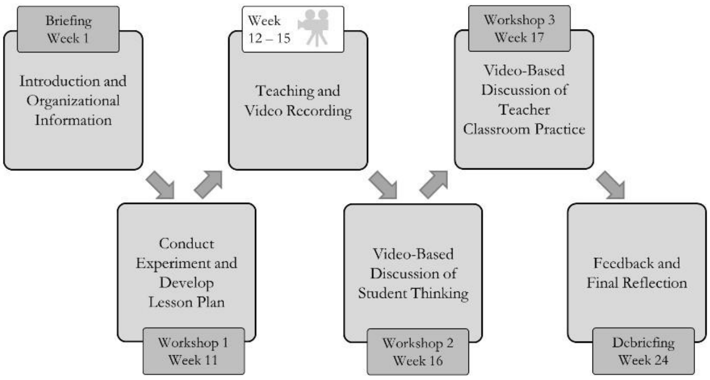

Fig. 2. Structure of the PD program on explicit instruction in experimentation (adapted from Koellner et al., 2007).

3.1.1 Briefing

The briefing is used to introduce the participating teachers to each other and to inform them about the goals of the PD program. With regard to possible reservations by the participants to show videos of their own classroom practice, it is important that the facilitator creates a pleasant and supportive environment throughout the whole PD program (Borko et al., 2008). In the briefing, the teachers are asked to explain what they expect from the PD program. The briefing also serves the purpose of providing the participants with organizational information about the accompanying research. To learn more about the teachers' current knowledge and beliefs about teaching experimentation, teachers' CK, PCK and beliefs are assessed as part of the briefing. Additionally, the teachers are informed about the pretest on classroom practice and student outcomes (Fig. 3).

3.1.2 Adaptation of Workshop 1

In Workshop 1 of the original PSC, teachers solve a "rich" mathematical problem (Borko et al., 2015). Such a problem provides a foundation for productive subject-related communication during the first workshop and is both challenging for teachers and appropriate for students with different levels of subject knowledge. Furthermore, it addresses multiple mathematical concepts and has various entry and exit points. Experimentation can also be described as a form of problem solving. The respective problems consist of developing hypotheses, generating data to test the hypotheses, as well as evaluating and coordinating evidence to draw a conclusion (Klahr & Dunbar, 1988; Mayer, 2007). Therefore, the teachers collaboratively conduct an experiment in groups in the adapted Workshop 1 (*collaborative community*). The teachers can choose from a number of prepared experiments that are appropriate for their own classes (e.g., measurement of the heartbeat of water fleas [*Daphnia pulex*] as a function of different water temperatures; Sannert & Hendel, 2021). These experiments are suitable to engage students in experimentation and to reflect on procedural knowledge (e.g., the control-of-variables strategy and its role in an experiment; OECD, 2017) and epistemic knowledge (e.g., how measurement error affects the degree of confidence in scientific knowledge; OECD, 2017). They have no clear "right answer" and are not too demanding in terms of the relevant content (Roberts & Reading, 2015). By conducting an experiment, the teachers switch to the students' perspective and discover possible ways to conduct the experiment and difficulties in conducting the experiment (*modeled inquiry*). After the experiment, the teachers reflect on their experiences and discuss which difficulties might occur for their students and which procedural and epistemic knowledge can best be taught explicitly with the experiment, in accordance with the national educational standards (*reflection, coherency*). This is first done in small groups and then in the whole group. As in the original Workshop 1, the teachers then plan a lesson for their own class by identifying learning goals, modifying the experiment as necessary, anticipating their students' solution strategies, and structuring their lesson (*developed lessons, transference*). The teachers begin to plan the lesson in small groups (*collaborative community*) and then finish the lesson planning by themselves. Because the teachers themselves might not have sufficient procedural and epistemic knowledge related to experimentation (Capps et al, 2016; Kamarainen et al., 2021), they are given appropriate supporting material in Workshop 1, which should help them to develop the necessary CK (*content knowledge*). In summary, Workshop 1 aims to primarily promote teachers's CK (procedural and epistemic knowledge related to experimentation). However, beyond that, PCK (KSU & KISR) is also addressed through *modeled inquiry* and *developed lessons*.

3.1.3 Teaching and Video Recording the Experiment

In the original PSC, the participants teach the planned lesson in their own class (Borko et al., 2011). The lessons are video recorded to capture the richness and complexity of classroom activities (Gaudin & Chaliès, 2015). Two cameras are used to film the teacher and the students when they are working in small-groups respectively (Borko et al., 2015). The PD facilitators select video clips, which then serve as a basis for analysis and reflection in Workshop 2 and Workshop 3. Video clips are selected based on the criteria of Borko et al. (e.g., "video clips foster reflection and discussion"; 2015, p. 16). Selecting video clips from the recordings of participants' classrooms is an adaptive approach, though less adaptive than others in which the participants bring their own videos (e.g., video clubs; Sherin & Van Es, 2009). However, through the selection of video clips and guiding questions, the facilitator can detect the needs and interests of the participants, while also considering the overall goal of the PD program (Borko et al., 2011). Furthermore, studies have shown that such scaffolds contribute to the quality of the analysis of videos (Hamel & Viau-Guay, 2019). Our PD program will not make any significant changes to these stages of the PSC.

3.1.4 Adaptation of Workshop 2 and Workshop 3

As in the original Workshops 2 and 3, in the adapted Workshops 2 and 3, the teachers share their experiences, supported by video clips of their teaching as well as by classroom documents (e.g., lesson plans and students' work). The video clips take a leading role in both workshops as a means for framing the teachers' collaborative exploration and for reflecting on central activities in the lesson taught (*reflection*), as well as for highlighting aspects of classroom practice that the teachers might not notice while teaching the lesson (Borko et al., 2011). The teachers' PCK is focused on in both workshops. The aim of this is not merely to help the teachers accumulate PCK but, instead, to support the teachers to activate and apply PCK in classroom practice (Borko et al., 2011).

The video-based analysis of the four or five videos (3-6 min each) first takes place individually on a laptop with headphones. In the video-based analysis, teachers are guided by three analysis tasks: (1) attend to relevant classroom situations in the video clips, (2) interpret these situations, and (3) generate possible alternatives for action on the basis of knowledge. We thereby refer to the concept of teacher noticing (Chan et al., 2021; Sherin & Van Es, 2009) and thus enable teachers to analyze the videos in a structured way. Watching their own video clips and video clips of their colleagues in similar situations allows the teachers to discover potential for improvement in their classroom practice and to understand that they face similar problems to other teachers; this can make it easier for them to change their classroom practice after identifying, interpreting, and discussing possible ways of teaching (Borko et al., 2008). Furthermore, watching videos of one's own teaching has been shown to be more activating and motivating than watching videos of unknown teachers (Seidel et al., 2011). After teachers have analyzed the videos individually, whole-group reflection begins (*collaborative community*). The PD facilitators prepare guiding questions as scaffolds to identify relevant classroom situations in the selected video clips and to structure the discussion about the video clips. Therefore, the selected video clips are used less to tell the teachers what to do next in the classroom and more as "springboards for analysis and discussions about teaching and learning" (Borko et al., 2011, p. 184). In order to structure the whole-group reflection and to engage teachers in discussion, facilitation moves from the framework for facilitating video-based PD are used (e.g., "pose general prompts to elicit participant ideas"; Van Es et al., 2014, p. 347). In Workshop 2, the PCK facet KSU is specifically addressed by taking *student thinking* into consideration. Video clips that show students struggling while engaging in experimentation or having inadequate conceptions with regard to procedural knowledge (e.g., the control-of-variables strategy) or epistemic knowledge (e.g., discussion about measuring errors) are of interest and will be selected. In Workshop 3, video clips are selected with the aim of focusing the discussions on the teachers' *classroom practice*, thereby addressing the PCK facet KISR. Video clips showing the teachers' adaptations of the experiment and different ways of integrating it into the lesson, as well as questions posed to the students to elicit procedural or epistemic knowledge related to experimentation, could potentially lead to fruitful discussions in this workshop. Given that student thinking and classroom practice are linked, the more salient of these two aspects is foregrounded in a workshop, while the other is placed in the background (Borko et al., 2015). Both workshops end with a final discussion in the whole group, in which the results are summarized and implications for future lesson planning and teaching are emphasized.

3.1.5 Debriefing

In the debriefing, the teachers report on their experiences with teaching experimentation (i.e., the integration of the PD program's content into their classroom) and the feedback they got from their students. The debriefing is also used to gain new ideas for instruction from the group or to discuss possible implementation difficulties. To see how teachers' professional competence has developed throughout the PD program, their CK, PCK, and beliefs are assessed as part of the debriefing (Fig. 3). Finally, the teachers are asked to give the PD facilitators feedback on the PD program. The debriefing ends with a reflection on what the teachers have learned over the whole PD program and on how it might have changed their classroom practice.

3.2 Sample

In total, 20 in-service biology teachers from the federal state of Schleswig-Holstein in Germany who are teaching biology in classes from Grade 5 to Grade 10 will be recruited to voluntarily take part in the PD program. All teachers of a school who are teaching biology from Grade 5 to Grade 10 will be recruited to take part collectively. Hence, in order to achieve a sample of 20 teachers, we aim to recruit teachers from four to five schools. Due to the sizes of the rooms available for the PD program and the group sizes recommended by Borko et al. (2015), each meeting (i.e., briefing, workshops, debriefing) will be carried out with six to 10 teachers, resulting in about two or three parallel PD programs being offered. Due to the special requirements of each school, the precise dates for the PD program will be planned individually with each school. As the comparison group for the student outcomes, we will use classes from the same grade level whose teachers are not taking part in the PD program. It is intended to match—as far as possible—the treatment group and the equally sized comparison group based on the control variables of the teachers (e.g., second subject, teacher experience). We will not include a comparison group for the teachers themselves because we assume that in-service teachers' professional competence is unlikely to develop in a linear and systematic way if they do not participate in an intervention (Kleickmann et al., 2013). The PD program will be conducted by the first author of this article, who will be supported by the second author. Both of them have jointly planned the PD activities as well as the accompanying study. Therefore, no specific training was conducted for the facilitators of the PD program.

3.3 Data Collection

In order to answer the first research question, we will measure teachers' knowledge and beliefs, classroom practice, and student outcomes in a pretest, a posttest, and a follow-up test (Fig. 3). As a rigorous research design should provide sufficient time between the PD program and the measurement of outcomes (Allen et al., 2011; Yoon et al., 2007), the posttests will be staggered after the PD program. Moreover, the follow-up test is planned to take place one year after the last workshop of the PD program. A period of one year is planned due to the fact that negative or positive experiences and feedback from the class could promote or inhibit the use of practices learned in the PD program. Furthermore, certain effects, especially on teachers' beliefs, may lag behind others such as teachers' knowledge or classroom practice (Guskey, 2002). To address the second research question, we will analyze qualitative data from the first research question as well as from the PD program. The data collection that is planned to measure the effects on teachers' knowledge and beliefs, classroom practice, and student outcomes is explained in more detail in the following sections.

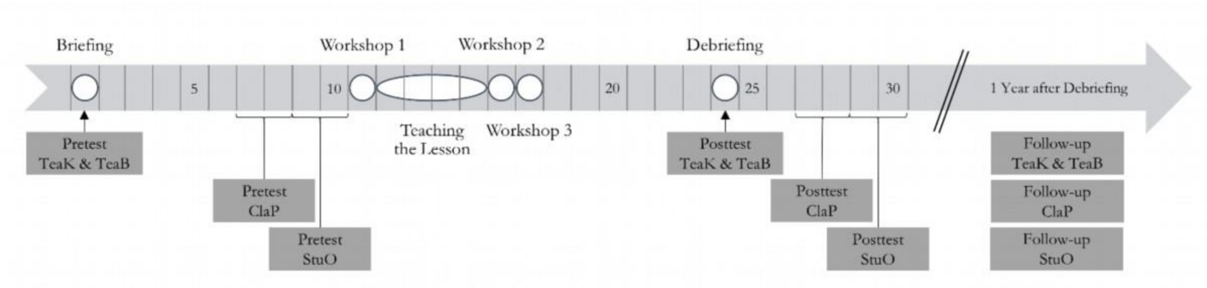

Fig. 3. Timeline of the study's procedure with the elements of the PD program in white circles/ellipses and planned data collections in dark gray boxes. Each vertical field represents one week.
TeaK = Teachers' Knowledge; TeaB = Teachers' Beliefs; ClaP = Classroom Practice; StuO = Student Outcomes

3.3.1 Effects on Teachers' Knowledge and Beliefs

For the assessment of the teachers' CK, 8 items from the multiple-choice instrument of Arnold (2015) have been selected. The items refer to procedural and epistemic knowledge that is related to experimentation, for example, knowledge about hypotheses (e.g., "What purpose do hypotheses serve in an experiment?"). The PCK necessary to teach experimentation will be assessed by using items from a multiple-choice instrument developed by Lieberei et al. (2023). The instrument includes 12 items that address the PCK facets KSU and KISR, as well as their interactions. The 8 items that focus on experimentation will be selected. Additionally, to gain broader and more qualitative insights into the teachers' PCK, we will use the *Content Representations* instrument (CoRe; Loughran et al., 2004). This instrument includes eight open-ended questions (e.g., "What difficulties and limitations do you expect when teaching this idea?"), which are usually used to investigate teachers' PCK on specific content (e.g., lactic acid; Barendsen & Henze, 2019). We will adapt the CoRe to elicit the PCK necessary to teach experimentation. Therefore, the teachers will be asked to provide written answers to eight questions for every "big idea" they have on teaching experimentation. (A big idea in our case could be, for example: "Use of an experiment to learn about dependent, independent, and control variables.") Regarding the teachers' beliefs about teaching experimentation, both object-related and self-related beliefs will be assessed (Hoy et al., 2006). To assess object-related beliefs, 55 items from a validated rating-scale instrument to measure beliefs about teaching experimentation will be selected and adapted (e.g., "fostering competencies in scientific inquiry requires extensive developing and summarizing phases, as well as extensive phases of fostering scientific content"; Petermann & Vorholzer, 2022). Teachers' self-efficacy beliefs regarding the teaching of experimentation will be captured using an 13-item adaptation of the STEBI (Riggs & Enochs, 1990), called the *Teaching Scientific Reasoning Efficacy Beliefs Instrument* (TSR-EBI; e.g., "I am continually finding better ways to teach scientific reasoning in my biology class."; Welter et al., 2023).

3.3.2 Effects on Classroom Practice

To see if classroom practice changes after teachers participate in the PD program, we will ask the teachers to show one typical lesson on experimentation before and after the PD program. Such a typical lesson has the advantage of being long enough to include key interactions between teachers and students, but short enough to be analyzable (Hiebert & Grouws, 2007). Classroom practice will be documented by means of a self-developed observation protocol, based on existing observation protocols published in the literature (Chen & Terada, 2021; Wee et al., 2007). The observation protocol will follow the criteria of a qualitative observation with a lower degree of complexity (Döring & Bortz, 2016). Accordingly, we have defined aspects on which the observation will focus (i.e., semi-structured observation). The observation focuses on (1) which scientific practices related to experimentation are addressed (e.g., asking questions, planning and carrying out investigations; Mayer, 2007), (2) the degree of openness of these practices (e.g., verification inquiry, structured inquiry, guided inquiry, open inquiry; Blanchard et al., 2010), and (3) which related procedural and epistemic knowledge (Krell et al., 2022) is made explicit to the students by the teacher. Each observation is conducted by two observers. To avoid observation errors, the observers will be trained in advance (e.g., trial observations based on video recordings; Döring & Bortz, 2016).

3.3.3 Effects on Student Outcomes

As discussed above, we plan to measure a broader conceptualization of student outcomes (Zeichner, 2005). Regarding the cognitive component of student outcomes, students' experimental competencies will be measured with 6 selected items from a multiple-choice instrument (Krell & Vierarm, 2016; Krell, 2018). The instrument was developed based on the model of *scientific discovery as dual search* (SDDS; Klahr & Dunbar, 1988). Furthermore, 8 selected items from a second instrument will be used to measure students' procedural and epistemic knowledge regarding experimentation (Arnold, 2015). With regard to the motivational component of student outcomes, students' motivation will be measured with items selected from Eccles and Wigfield (1995), Pugh et al. (2010), and Simpkins et al. (2006). These items include the factors interest (3 items), self-concept (3 items), and value (8 items).

3.3.4 Teachers' Content-Related Learning Processes

To identify learning pathways and learning obstacles in the teachers' implementation of experimentation, we will analyze qualitative data from the observation protocol (pre- and posttest), transcripts from the videotaped lessons, lesson plans, and the CoRe instrument (pre- and posttest). Furthermore, the teachers will be asked to provide us with their lesson plans up to the follow-up test showing how they engaged their students in experimentation during and after the PD program.

3.4 Data Analysis

In the following, the methods that we plan to use to analyze the data for each of the two research questions are explained.

3.4.1 Effects of the PD Program

Research question one asks about the effects of participating in the PD program on the participating teachers' CK, PCK, and beliefs, on the quality of their classroom practice, and on student outcomes related to scientific reasoning. In order to address the research question, we will carry out the first evaluation after the posttest to evaluate the immediate effectiveness of the PD program. To measure the stability of the effects, the values obtained for the respective measures in the follow-up test will be considered as well. All qualitative data from the CoRe instrument and classroom observations will be analyzed by means of qualitative content analysis (Schreier, 2012) considering quality ensuring procedures (e.g., checking intercoder agreement as a measure of coding objectivity; Göhner & Krell, 2020). We will analyze the data using theoretically derived categories and inductively refine and specify the categories based on the data. For instance, for the analysis of classroom observation, the categories (1) scientific practices (e.g., asking questions, planning and carrying out investigations; Mayer, 2007), (2) degree of openness (e.g., verification inquiry, structured inquiry, guided inquiry, open inquiry; Blanchard et al., 2010), and (3) related procedural and epistemic knowledge (Krell et al., 2022) will be used as initial categories. In addition, we will transform the qualitative data into quantitative data to be used in the quantitative data analysis. All quantitative data, including the comparison group for student outcomes, will be analyzed by means of appropriate parametric or nonparametric tests for dependent group comparisons (e.g., paired *t*-test, ANOVA with repeated measures) to test for significant differences in the variables between pre-, post- and follow-up test. Furthermore, we will compare possible differences between experimental group and control group in terms of student outcomes.

3.4.2 Teachers' Content-Related Learning Processes

Research question two asks about the content-related learning pathways of teachers during the PD program, as well as about obstacles preventing implementation in classroom practice. We will use qualitative content analysis to analyze the data obtained with the observation protocol, transcripts from the video-recorded lessons, lesson plans, and data from the CoRe instrument. Again, we will analyze the data using theoretically derived categories and inductively specify the categories based on the data. With these data, we plan to identify starting points for how to professionalize the teaching of experimentation and also to identify obstacles preventing the integration of the content of the PD program into classroom practice. With the findings we gain from these data, we aim to specify the design feature of *content focus* in a more content-specific manner.

4 Discussion and Conclusions

With our study, we aim to obtain evidence on the hitherto rarely researched question of how a PD program affects teachers' knowledge and beliefs, the quality of classroom practice, and student learning outcomes related to scientific reasoning. We therefore designed a PD program on explicit instruction in experimentation. The PD program is an adaptation of the PSC (Koellner et al., 2007) and incorporates nine design features of effective PD programs for inquiry-based teaching (Capps et al., 2012; Ramnarain et al., 2022): *content knowledge, developed lessons, modeled inquiry, reflection, coherency, transference, total time, extended support*, and *collaborative community*. We argue that these design features can be considered as specifications of the generic design features that PD programs should have in order to be effective (Desimone, 2009). Following the framework that we developed to study the effects of the PD program, we will measure teachers' knowledge (CK and PCK), teachers' beliefs, and their self-efficacy beliefs related to teaching experimentation. In addition, we will use classroom observation to obtain direct information about changes in classroom practice (Jacobs et al., 2014). As student outcomes, cognitive components (e.g., procedural and epistemic knowledge) and motivational components will be measured. In order to answer the question of which content-related learning pathways teachers typically take during the PD program and what obstacles prevent implementation in classroom practice, we will analyze qualitative data from the observation protocol, transcripts from the videotaped lessons, lesson plans, and the CoRe instrument. The results will contribute to a more detailed description of the design feature *content knowledge* regarding PD programs that intend to improve students' scientific reasoning competencies through experimentation.

Although the structure with several cycles is an important characteristic of the original PSC (Borko et al., 2015), our adapted PD program does not include several cycles but instead has a linear structure. We made this decision for two reasons: First, it is the first implementation of the PD program and should therefore be examined more closely before a second cycle is started. Second, we believe that several cycles may prevent teachers from participating in the PD program, because there is usually no extra time planned for in-service teachers' PD in Germany (Richter et al., 2020). A result of this may be that many teachers invest little time in PD programs and are used to one-shot workshops instead of continuous PD programs with longer durations. The median duration of the PD programs attended in Schleswig-Holstein is 6.5 hr (Richter & Schellenbach-Zell, 2016). For these reasons, we decided to operate at the lower limit of the duration that is considered necessary for a PD program to be effective (Yoon et al., 2007) and to not offer a subsequent second cycle.

In our data collection, the posttests will be staggered after the PD program. In line with Yoon et al. (2007), we expect to find effects on teachers' knowledge and beliefs first, followed by changes in classroom practice, and, finally, improved student outcomes. The possibilities for teachers to implement the content of the PD program in classroom practice are limited for curricular reasons. Furthermore, the process of implementation takes time beyond the PD program (Guskey, 2002). It is recommended that there is a sufficiently long interval between completion of the intervention and the evaluation (Allen et al., 2011; Lipowsky & Rzejak, 2019). Therefore, we plan to conduct a follow-up evaluation a full year after the PD program. The data collection is planned based on the expectation that 20 teachers will participate in the PD program. If fewer teachers participate, we will choose more qualitative methods for the data collection and data analysis. If more teachers are recruited, we will consider further statistical methods for the data analysis on the effects on student outcomes, such as structural equation modeling and hierarchical linear modeling (Cheong & MacKinnon, 2012; Nezlek et al., 2006).

For future research, it may be of interest to adapt the PD program to other scientific methods (e.g., modeling) and also to see how the adapted PD program works for further types of schools (e.g., in inclusive settings).

Acknowledgements

The authors would like to thank Gráinne Newcombe for English language editing.

References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045), 1034–1037. https://doi.org/10.1126/science.1207998

Arnold, J. (2015). *Die Wirksamkeit von Lernunterstützungen beim forschenden Lernen: Eine Interventionsstudie zur Förderung des wissenschaftlichen Denkens in der gymnasialen Oberstufe*. Logos. http://hdl.handle.net/11654/25738

Bandura, A. (1997). *Self-efficacy: The exercise of control.* Freeman.

Barendsen, E., & Henze, I. (2019). Relating teacher PCK and teacher practice using classroom observation. *Research in Science Education*, *49*(5), 1141–1175. https://doi.org/10.1007/s11165-017-9637-z

Barni, D., Danioni, F., & Benevene, P. (2019). Teachers' self-efficacy: The role of personal values and motivations for teaching. *Frontiers in psychology*, *10*, 1645. https://doi.org/10.3389/fpsyg.2019.01645

Baumert, J., & Kunter, M. (2013a). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Kraus & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 25–48). Springer. https://doi.org/10.1007/978-1-4614-5149-5_2

Baumert, J., & Kunter, M. (2013b). The effect of content knowledge and pedagogical content knowledge on instructional quality and student achievement. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Kraus & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 175–205). Springer. https://doi.org/10.1007/978-1-4614-5149-5_9

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science education, 94*(4), 577–616. https://doi.org/10.1002/sce.20390

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and teacher education*, *24*(2), 417–436. https://doi.org/10.1016/j.tate.2006.11.012

Borko, H., Koellner, K., Jacobs, J., & Seago, N. (2011). Using video representations of teaching in practice-based professional development programs. *ZDM Mathematics Education, 43*, 175–187. https://doi.org/10.1007/s11858-010-0302-5

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. In E. Baker, B. McGaw & P. Peterson (Eds.), *International Encyclopedia of Education* (pp. 548–555). Elsevier. https://doi.org/10.1016/B978-0-08-044894-7.00654-0

Borko, H., Jacobs, J., Koellner, K., & Swackhamer, L. E. (2015). *Mathematics professional development: Improving teaching using the problem-solving cycle and leadership preparation models*. Teachers College Press. https://doi.org/10.1080/10986060701360944

Buehl, M. M., & Beck, J. S. (2015). The relationship between teachers' beliefs and teachers' practices. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 66–84). Routledge. https://doi.org/10.4324/9780203108437-11

Büssing, A. G., Dietz, C., Grahmann, M., Klein, H. P., & Möller, A. (2022). *Characteristics and Predictors of Scientific Practices: A Large-scale Study* [online preprint]. https://doi.org/10.31219/osf.io/yfmk8

Cairns, D., & Areepattamannil, S. (2019). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in science education*, *49*(1), 1–23. https://doi.org/10.1007/s11165-017-9639-x

Capps, D. K., & Crawford, B. A. (2013). Inquiry-based instruction and teaching about nature of science: Are they happening? *Journal of Science Teacher Education, 24*(3), 497–526. https://doi.org/10.1007/s10972-012-9314-z

Capps, D. K., Crawford, B. A., & Constas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. *Journal of science teacher education*, *23*(3), 291–318. https://doi.org/10.1007/s10972-012-9275-2

Capps, D. K., Shemwell, J. T., & Young, A. M. (2016). Over reported and misunderstood? A study of teachers' reported enactment and knowledge of inquiry-based science teaching. *International Journal of Science Education*, *38*(6), 934–959. https://doi.org/10.1080/09500693.2016.1173261

Chan, K. K. H., Xu, L., Cooper, R., Berry, A., & van Driel, J. H. (2021). Teacher noticing in science education: do you see what I see? *Studies in Science Education, 57*(1), 1–44. https://doi.org/10.1080/03057267.2020.1755803

Chen, Y. C., & Terada, T. (2021). Development and validation of an observation-based protocol to measure the eight scientific practices of the next generation science standards in K-12 science classrooms. *Journal of Research in Science Teaching*, *58*(10), 1489–1526. https://doi.org/10.1002/tea.21716

Cheong, J., & MacKinnon, D. (2012). Mediation/indirect effects in structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 417–435). Guilford Press.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective Teacher Professional Development*. Learning Policy Institute. https://learningpolicyinstitute.org/product/teacher-prof-dev

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

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society and Education, 7*(3), 252–263.

Diamond, B. S., Maerten-Rivera, J., Rohrer, R. E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 results. *Journal of Research in Science Teaching, 51*(5), 635–658. https://doi.org/10.1002/tea.21148

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer. https://doi.org/10.1007/978-3-642-41089-5

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215–225. https://doi.org/10.1177/0146167295213003

Enzingmüller, C., & Prechtl, H. (2021). Constructing graphs in biology class: Secondary biology teachers' beliefs, motivation, and self-reported practices. *International Journal of Science and Mathematics Education*, *19*(1), 1–19. https://doi.org/10.1007/s10763-019-09975-2

Fischer, C., Fishman, B., Dede, C., Eisenkraft, A., Frumin, K., Foster, B., & McCoy, A. (2018). Investigating relationships between school context, teacher professional development, teaching practices, and student achievement in response to a nationwide science reform. *Teaching & Teacher Education, 72,* 107–121. https://doi.org/10.1016/j.tate.2018.02.011

Fives, H., & Gill, M. G. (2015). *International handbook of research on teachers' beliefs.* Routledge.

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

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., & Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011–4024). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://eric.ed.gov/?id=ED519922

Gess-Newsome, J., Taylor, J. A., Carlson, J., Gardner, A. L., Wilson, C. D., & Stuhlsatz, M. A. (2019). Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education*, *41*(7), 944–963. https://doi.org/10.1080/09500693.2016.1265158

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American educational research journal*, *37*(2), 479–507. https://doi.org/10.3102/00028312037002479

Göhner, M., & Krell, M. (2020). Qualitative Inhaltsanalyse in naturwissenschaftsdidaktischer Forschung unter Berücksichtigung von Gütekriterien: Ein Review [Qualitative content analysis in science education research under the consideration of quality criteria]. *Zeitschrift für Didaktik der Naturwissenschaften, 26*(1), 207–225. https://doi.org/10.1007/s40573-020-00111-0

Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. (2014). Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of mathematics teacher education*, *17*(1), 5–36. https://doi.org/10.1007/s10857-013-9245-4

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

Hamel, C., & Viau-Guay, A. (2019). Using video to support teachers' reflective practice: A literature review. *Cogent Education*, *6*(1), 1–14. https://doi.org/10.1080/2331186X.2019.1673689

Hameleers, M., & Van der Meer, T. G. (2021). The scientists have betrayed us! The effects of anti-science communication on negative perceptions toward the scientific community. *International Journal of Press/Politics*, *26*(1), 46–68.

Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). IAP.

Hoy, A. W., Davis, H., & Pape, S. J. (2006). Teacher Knowledge and Beliefs. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 715–737). Lawrence Erlbaum Associates Publishers.

Jacobs, J., Koellner, K., John, T., & King, C. D. (2014). The process of instructional change: Insights from the Problem-Solving Cycle. In Y. Li, E. A. Silver & S. Li (Eds.), *Transforming mathematics instruction* (pp. 335–354). Springer.

Kamarainen, A., Grotzer, T., Thompson, M., Sabey, D., & Haag, B. (2021). Teacher views of experimentation in ecosystem science. *Journal of Biological Education*, 1-20. https://doi.org/10.1080/00219266.2021.1933130

Kelcey, B., Hill, H. C., & Chin, M. (2019). Teacher mathematical knowledge, instructional quality, and student outcomes: A multilevel quantile mediation analysis. *School Effectiveness and School Improvement, 30*, 398–431. https://doi.org/10.1080/09243453.2019.1570944

Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, *54*(5), 586–614. https://doi.org/10.1002/tea.21378

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

Kind, P. E. R., & Osborne, J. (2016). Styles of scientific reasoning: A cultural rationale for science education? *Science education*, *101*(1), 8–31. https://doi.org/10.1002/sce.21251

Kind, V., Park, S., & Chan, K. K. H. (2022). Science teacher professional knowledge and its relationship to high-quality science instruction. In J. A. Luft & M. G. Jones (Eds.), *Handbook of Research on Science Teacher Education* (pp. 329–339). Routledge.

Kirkpatrick, D.L., & Kirkpatrick, J. D. (1994). *Evaluating training programs. The four levels*. Berrett-Koehler.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 54*, 1–48. https://doi.org/10.1207/s15516709cog1201_1

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of teacher education*, *64*(1), 90–106. https://doi.org/10.1177/0022487112460398

KMK (2020). *Bildungsstandards im Fach Biologie für die Allgemeine Hochschulreife [Educational Standards in Biology for the Higher Education Entrance Qualification]*. Wolters Kluwer.

Koellner, K., Jacobs, J., Borko, H., Schneider, C., Pittman, M. E., Eiteljorg, E., ... & Frykholm, J. (2007). The problem-solving cycle: A model to support the development of teachers' professional knowledge. *Mathematical thinking and learning*, *9*(3), 273–303. https://doi.org/10.1080/10986060701360944

Koellner, K., Jacobs, J., Borko, H., Roberts, S., & Schneider, C. (2011). Professional development to support students' algebraic reasoning: An example from the Problem-Solving Cycle Model. In C. Jinfa & E. Knuth (Eds.), *Early algebraization* (pp. 429–452). Springer. https://doi.org/10.1007/978-3-642-17735-4_23

Krell, M. (2018). Schwierigkeitserzeugende Aufgabenmerkmale bei Multiple-Choice-Aufgaben zur Experimentierkompetenz im Biologieunterricht: Eine Replikationsstudie [Difficulty-generating task characteristics in multiple-choice tasks for experimental competence in biology teaching: A replication study]. *Zeitschrift für Didaktik der Naturwissenschaften, 24*(1), 1–15. https://doi.org/10.1007/s40573-017-0069-0

Krell, M., & Vierarm, A. (2016). Analyse schwierigkeitserzeugender Aufgabenmerkmale bei einem Multiple-Choice-Test zum Experimentieren [Analysis of difficulty generating characteristics of a multiple choice test assessing competencies in experimentation]. In U. Gebhard & M. Hammann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* (pp. 283–298). Studienverlag.

Krell, M., Vorholzer, A., & Nehring, A. (2022). Scientific reasoning in science education: From global measures to fine-grained descriptions of students' competencies. *Education Sciences*, *12*(2), 97. https://doi.org/10.3390/educsci12020097

Kulgemeyer, C., & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, *55*(10), 1393–1418. https://doi.org/10.1002/tea.21457

Lederman, N. G., & Lederman, J. S. (2019). Teaching and learning of nature of scientific knowledge and scientific inquiry: building capacity through systematic research-based professional development. *Journal of Science Teacher Education*, *30*(7), 737–762. https://doi.org/10.1080/1046560X.2019.1625572

Lieberei, T., Welter, V., Großmann, L., & Krell, M. (2023). *Findings from the expert-novice paradigm on differential response behavior among multiple-choice items of a PCK test* [Manuscript submitted for publication].

Lipowsky, F., & Rzejak, D. (2019). Was macht Fortbildungen für Lehrkräfte erfolgreich? Ein Update [What makes PD programs for teachers effective? – An update]. In B. Groot-Wilken & R. Koerber (Eds.), *Nachhaltige Professionalisierung für Lehrerinnen und Lehrer: Ideen, Entwicklungen, Konzepte* (pp. 15–56). wbv Publikation.

Lotter, C., Carnes, N., Marshall, J. C., Hoppmann, R., Kiernan, D. A., Barth, S. G., & Smith, C. (2020). Teachers' content knowledge, beliefs, and practice after a project-based professional development program with ultrasound scanning. *Journal of Science Teacher Education*, *31*(3), 311–334. https://doi.org/10.1080/1046560X.2019.1705535

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching, 41*(4), 370–391. https://doi.org/10.1002/tea.20007

Lumpe, A., Czerniak, C., Haney, J., & Beltyukova, S. (2012). Beliefs about teaching science: The relationship between elementary teachers' participation in professional development and student achievement. *International journal of science education*, *34*(2), 153–166. https://doi.org/10.1080/09500693.2010.551222

Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen [Knowledge acquisition as scientific problem solving]. In D. Krüger & H. Vogt (Eds.), *Theorien in Der Biologiedidaktischen Forschung [Theories in Biology Education Research]*, (pp. 177–186). Springer.

Mohamadi, F. S., & Asadzadeh, H. (2012). Testing the mediating role of teachers' self-efficacy beliefs in the relationship between sources of efficacy information and students achievement. *Asia Pacific Education Review*, *13*(3), 427–433. https://doi.org/10.1007/s12564-011-9203-8

Nezlek, J., Schröder-Abé, M., & Schütz, A. (2006). Mehrebenenanalysen in der psychologischen Forschung: Vorteile und Möglichkeiten der Mehrebenenmodellierung mit Zufallskoeffizienten. *Psychologische Rundschau, 57*, 213–223. https://doi.org/10.1026/0033-3042.57.4.213

NRC (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press.

OECD (2017). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving (Revised Edition)*. OECD Publishing.

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education, 25*(2), 177–196. https://doi.org/10.1007/s10972-014-9384-1

Osborne, J. et al., (2022). *Science education in an age of misinformation*. Stanford University. https://policycommons.net/artifacts/2434623/science_education_in_an_age_of_misinformation/3456215

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of educational research*, *62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Park, S., & Chen, Y. C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of research in science teaching*, *49*(7), 922–941. https://doi.org/10.1002/tea.21022

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American educational research journal, 44*(4), 921–958. https://doi.org/10.3102/0002831207308221

Petermann, V., & Vorholzer, A. (2022). Relationship between beliefs of teachers about and their use of explicit instruction when fostering students' scientific inquiry competencies. *Education Sciences*, *12*(9), 593. https://doi.org/10.3390/educsci12090593

Prediger, S., Leuders, T., & Rösken-Winter, B. (2017). Drei-Tetraeder-Modell der gegenstandsbezogenen Professionalisierungsforschung: Fachspezifische Verknüpfung von Design und Forschung. *Jahrbuch für allgemeine Didaktik*, 159–177.

Prediger, S., Schnell, S., & Rösike, K.-A. (2016). Design research with a focus on content-specific professionalisation processes: The case of noticing students' potentials. In S. Zehetmeier, B. Rösken-Winter, D. Potari, & M. Ribeiro (Eds.), *Proceedings of the third ERME topic conference on mathematics teaching, resources and teacher professional development* (pp. 96–105). Humboldt University/HAL Archive.

Pugh, K. J., Linnenbrink-Garcia, L., Koskey, K. L., Stewart, V. C., & Manzey, C. (2010). Motivation, learning, and transformative experience: A study of deep engagement in science. *Science Education, 94*(1), 1–28. https://doi.org/10.1002/sce.20344

Ramnarain, U., Capps, D., & Hsu, Y. S. (2022). Professional development of science teachers for inquiry instruction. In J. A. Luft & M. G. Jones (Eds.), *Handbook of Research on Science Teacher Education* (pp. 273–286). Routledge.

Richter, D., & Schellenbach-Zell, J. (2016). Fort- und Weiterbildung von Lehrkräften in Schleswig-Holstein: Ergebnisse einer Befragung im Jahr 2016 [Professional development of teachers in Schleswig-Holstein: Results of a survey in 2016]. Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein.

Richter, E., Marx, A., Huang, Y., & Richter, D. (2020). Zeiten zum beruflichen Lernen: Eine empirische Untersuchung zum Zeitpunkt und der Dauer von Fortbildungsangeboten für Lehrkräfte. *Zeitschrift für Erziehungswissenschaft*, *23*(1), 145–173. https://doi.org/10.1007/s11618-019-00924-x

Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*(6), 625–637. https://doi.org/10.1002/sce.3730740605

Roberts, R., & Reading, C. (2015). The practical work challenge: Incorporating the explicit teaching of evidence in subject content. *School Science Review*, *357*, 31–39.

Roesken-Winter, B., Stahnke, R., Prediger, S., & Gasteiger, H. (2021). Towards a research base for implementation strategies addressing mathematics teachers and facilitators. *ZDM Mathematics Education*, *53*(5), 1007–1019. https://doi.org/10.1007/s11858-021-01220-x

Rychen, D. S., & Salganik, L. H. (2003). A holistic model of competence. In D. S. Rychen & L. H. Salganik (Eds.), *Key competencies for a successful life and a well-functioning society* (pp. 41–62). Hogrefe and Huber Publishers.

Sannert, R., & Hendel, B. (2021). Plastik in Tieren – Plastik in uns? Einflüsse von Mikrostoffen auf Daphnien mit Counter-App messen. *Digital unterrichten BIOLOGIE, 2*(1), 10–11.

Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and teacher education*, *27*(2), 259–267. https://doi.org/10.1016/j.tate.2010.08.009

Sharon, A. J., & Baram-Tsabari, A. (2020). Can science literacy help individuals identify misinformation in everyday life? *Science Education, 104*(5), 873–894. https://doi.org/10.1002/sce.21581

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

Sherin, M., & Van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of teacher education, 60*(1), 20–37. https://doi.org/10.1177/0022487108328155

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42*(1), 70–83. https://doi.org/10.1037/0012-1649.42.1.70

Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: a systematic review of empirical mathematics education research. *ZDM Mathematics Education*, *48*(1), 1–27. https://doi.org/10.1007/s11858-016-0775-y

Sztajn, P. (2011). Standards for reporting mathematics professional development in research studies. *Journal for Research in Mathematics Education, 42*(3), 220–236. https://doi.org/10.5951/jresematheduc.42.3.0220

Van Driel, J. H., Meirink, J. A., van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: a review of design features and quality of research. *Studies in science education, 48*(2), 129–160. https://doi.org/10.1080/03057267.2012.738020

Van Es, E. A., Tunney, J., Goldsmith, L. T., & Seago, N. (2014). A framework for the facilitation of teachers' analysis of video. *Journal of teacher education*, *65*(4), 340–356. https://doi.org/10.1177/0022487114534266

Vorholzer, A., von Aufschnaiter, C., Boone, W.J. (2020). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education, 50*, 333–359. https://doi.org/10.1007/s11165-018-9691-1

Wagensveld, B., Segers, E., Kleemans, T., & Verhoeven, L. (2015). Child predictors of learning to control variables via instruction or self-discovery. *Instructional Science*, *43*(3), 365–379. https://doi.org/10.1007/s11251-014-9334-5

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competences* (pp. 45–65). Hogrefe and Huber Publishers.

Wee, B., Shepardson, D., Fast, J., & Harbor, J. (2007). Teaching and learning about inquiry: Insights and challenges in professional development. *Journal of science teacher education*, *18*(1), 63–89. https://doi.org/10.1007/s10972-006-9031-6

Welter, V., Dawborn-Gundlach, M., Großmann, L., & Krell, M. (2023). *Adapting a self-efficacy scale to the task of teaching scientific reasoning* [Manuscript submitted for publication].

Yang, Y., Liu, X., & Gardella Jr, J. A. (2020). Effects of a professional development program on science teacher knowledge and practice, and student understanding of interdisciplinary science concepts. *Journal of Research in Science Teaching*, *57*(7), 1028–1057. https://doi.org/10.1002/tea.21620

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. http://ies.ed.gov/ncee/edlabs

Zeichner, K. M. (2005). A research agenda for teacher education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 737–759). American Educational Research Association.