Integrating or Not-Integrating—That is the Question. Effects of Integrated Instruction on the Development of Pre-Service Biology Teachers' Professional Knowledge

Maria Kramer *[†], Christian Förtsch[†] and Birgit J. Neuhaus[†]

Biology Education, Department I, Faculty of Biology, LMU Munich, Munich, Germany

For successful classroom instruction, teachers require a well-founded knowledge base consisting of the three knowledge facets pedagogical-psychological knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK). However, there is not yet clarity about the circumstances and instructional pathways through which teachers can best develop these knowledge facets. In an experimental study (N = 118 pre-service biology teachers), we investigated the effects of separated instruction (knowledge facets were treated successively without linking) or integrated instruction (knowledge facets were presented in an interrelated way) on PK, CK, PCK and the application of PCK in a video-based assessment tool in comparison to a control group (receiving no instruction). Both pathways of instruction were provided by a lecturer on the curricular topic of senses and sensory organs, exemplified for the topic skin. Results point to the effectiveness of both ways of instruction in terms of knowledge increases for CK and PCK. In addition, working with the video-based assessment tool may have had an additional effect on PCK. No effects for PK could be found, possibly due to a ceiling effect. Moreover, there was no effect of the intervention on the application of PCK. However, tendencies in descriptive results indicating a possible advantage concerning separated or integrated instruction with regard to CK and PCK are discussed. Overall, our results indicate that the use of video-based tools can complement instructional approaches to knowledge acquisition.

Keywords: teacher education-prospective teachers, pedagogical content knowledge, instructional design, teacher professional knowledge, integrated instruction, biology education

INTRODUCTION

Teacher education programs support pre-service teachers in acquiring professional knowledge that is fundamental for high-quality instruction (Baumert et al., 2010; Keller et al., 2017). Direct instructional guidance is one way to support knowledge acquisition and is therefore an important element of lectures or seminars that pre-service teachers attend at university. Direct instructional guidance can be understood as "providing information that fully explains the concepts and procedures that students are required to learn as well as learning strategy support that is compatible with human cognitive architecture" (Kirschner et al., 2006, p. 75). Concepts and procedures that pre-service teachers have to know about for powerful teaching mainly relate to

three different facets of professional knowledge. These facets cover knowledge of subject-specific core ideas (content knowledge or CK), knowledge of subject-specific strategies to make subject-specific core ideas and content accessible for students (pedagogical content knowledge or PCK), and knowledge of general pedagogical-psychological principles and methods (pedagogical-psychological knowledge or PK) (Shulman, 1987; Voss et al., 2011; Baumert and Kunter, 2013a). Although there is evidence that linking and cross-referencing between knowledge facets is crucial for their retrievability and applicability, the knowledge facets are mainly addressed within separate university courses that rarely connect content and pedagogy (Renkl et al., 1996; Harr et al., 2015; König et al., 2018; Tröbst et al., 2019). For example, pre-service biology teacher attend courses in pedagogy, in which they are instructed in general teaching methods and strategies for classroom management (Voss et al., 2011). In courses of the discipline biology, pre-service teachers then acquire knowledge about specific biological topics (e.g., about the skin and its structure); whereas within didactical courses they get to know, for example, core ideas such as *structure and function*, and how to implement them in biology instruction (Sekretariat der Ständigen Konferenz der Kultusminister der Länderin der Bundesrepublik Deutschland, 2005; National Research Council, 2012). Pre-service teachers also get to know strategies for dealing with student ideas and planning concept-oriented lessons that can be considered as part of teachers' PCK. The integration of information from the three knowledge facets is largely the pre-service teachers' own task. In other words, how to deal with student ideas about the skin and its structure, and how to use scientific core ideas to foster students' understanding of the specific content is rarely explicitly addressed. Since this task creates considerable difficulties and is hardly achieved successfully, researchers call for an integration of knowledge facets within teacher education, which makes the transformation into effective classroom instruction more likely (Renkl et al., 1996; Ball, 2000; Kleickmann et al., 2017; Tröbst et al., 2019). Studies focusing on the effects of integrated instruction in university courses, which consider all three knowledge facets for instructional input and as outcome measure, have hardly been conducted so far. To decide when and under which circumstances integration is appropriate, more empirical data from different domains is needed.

Therefore, the present study experimentally compared the effects of instructing all three knowledge facets, PK, CK, and PCK, in a separated or integrated way. The separated condition treats aspects of PK, CK, and PCK successively without linking and cross-referencing content, as it is usually done in university teacher education (Harr et al., 2014; König et al., 2018). In the integrated condition, a lesson planning model was used to structure instruction. For each phase of the planning model, corresponding aspects of all three knowledge facets were then presented in an interrelated way. Both conditions received direct instructional guidance of the knowledge facets through a lecturer. Thus, the chosen instructional approach allows at the same time a very practical investigation of the

effects of integrated instruction since teacher training programs often include courses guided by a lecturer (Tröbst et al., 2019). Therefore, especially for designing teacher education programs, the present study is of great practical use as well.

Professional Knowledge as Part of Teachers' Professional Competence

Teachers' professional competence describes how teachers, depending on cognitive and affective-motivational dispositions, apply specific skills in specific situations to inform their actions. The understanding of competence described therein is represented in the competence-as-a-continuum model (Blömeke et al., 2015), which can be applied to different contexts such as lesson planning, instructing, reflecting on instruction, or diagnosing. The integration of professional knowledge facets that are part of teachers' cognitive dispositions counts as an important key process within the varying contexts in order to act effectively (Brunner et al., 2013; Heitzmann et al., 2019).

The importance of teachers' professional knowledge for instructional quality and student outcomes is empirically well-proven (Baumert et al., 2010; Fischer et al., 2012; Kunter et al., 2013; Förtsch et al., 2016). With regard to the subject and specific contextual and situational demands, different facets of professional knowledge have been distinguished on the basis of Shulman (1987). The common ground is that teachers need pedagogical-psychological knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK) during instruction (Shulman, 1986, 1987; Baumert et al., 2010; Voss and Kunter, 2013). PK is considered generic and independent from a specific subject, and has been conceptualized as knowledge about classroom processes and students' heterogeneity (Voss et al., 2011) or as knowledge about generic theories and methods of instruction and learning as well as of classroom management (König et al., 2014).

Furthermore, when teaching a subject, teachers need knowledge of subject matter. Professional knowledge referring to the understanding of subject-specific methods and core concepts is called content knowledge (CK) (Ball et al., 2008; Gess-Newsome, 2015). The knowledge that is necessary to make this content available for a particular group of students is referred to as pedagogical content knowledge (PCK), including aspects of content and pedagogy (Shulman, 1987; Gess-Newsome, 2015). Although conceptualizations of PCK differ, they all emphasized knowledge about subject-specific instruction and a student-related perspective as essential. Therefore, important components of PCK are knowledge about student (mis)conceptions as well as knowledge about subject-specific structures of instruction and corresponding teaching strategies (Park and Oliver, 2008; Depaepe et al., 2013; Schmelzing et al., 2013). While there are conceptualizations in which both CK and PCK are included as subject-specific knowledge for teaching (e.g., Ball et al., 2008; Hill et al., 2008; Kersting et al., 2010), other researchers developed instruments for measuring knowledge facets separately (e.g., Jüttner et al., 2013). Results from the COACTIV project

(*Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*) led to the conclusion that CK and PCK exist as two overlapping but distinct facets (Krauss et al., 2013). However, within the ProwiN project (*Professional Knowledge of Teachers in Science*) CK and PCK did not correlate significantly, and thus, could be measured independently (Förtsch et al., 2016).

In addition, different types of the knowledge facets are distinguished that can be referred to as declarative (i.e., knowledge related to facts, terms and principles) and action-related knowledge (i.e., knowledge about actions, manipulations, or procedures, as well as knowledge about when and why to apply these procedures in order to foster student learning) (Förtsch et al., 2018). When measuring professional knowledge, declarative knowledge and action-related knowledge were mostly addressed in the form of paper-pencil tests, including short answer or multiple-choice formats (Harr et al., 2014; Förtsch et al., 2018). While paper-pencil tests, mainly used within quantitative approaches, are connected to a more cognitive perspective on professional knowledge, there are also approaches to study teachers' professional knowledge from a situated perspective (Depaepe et al., 2013). Within the situated perspective, professional knowledge can also be understood as dynamic knowledge in action (Alonzo and Kim, 2016) or integrated knowledge that teachers apply to observe and evaluate classroom instruction and to identify possible challenges (Kersting et al., 2010; Seidel et al., 2013). Consequently, knowledge should be captured within specific contexts that are closer to practice. That can be done, for example, with text or video vignettes in which authentic teaching situations are presented (Cauet et al., 2015; Hoth et al., 2018). Observations or subsequent interviews can then provide insights into the teachers' knowledge. However, this is usually realized within qualitative approaches (cf. Alonzo and Kim, 2018). The distinction between cognitive and situated perspectives is also displayed within recent models such as the Refined Consensus Model that depicts teachers' professional knowledge in terms of the facets PK, CK, and PCK within the science classroom, but also takes different realms of PCK into account (Carlson et al., 2019). One realm of PCK describes a kind of canonical knowledge that professionals of a discipline share and that is taught in university courses (*collective* PCK). This form differs, however, from the personal knowledge (*personal* PCK) that each pre-service teacher holds. Researchers assume that *personal* PCK develops based on the knowledge that is taught and the persons' individual experiences within the classroom. The third form of PCK refers to the knowledge that "teachers draw on in the moment of action, where the action may include planning, teaching, or reflecting on teaching" (Alonzo et al., 2019, *p.* 273) and is therefore referred to as *enacted* PCK. Whereas *collective* and *personal* PCK represent more static forms of PCK, and thus, are similar to the cognitive perspective, *enacted* PCK is more adaptive and connected to a specific classroom situation. Thus, *enacted* PCK is related to the study of knowledge and skills from the situated perspective (Alonzo et al., 2019). Eventually, both perspectives are important as they complement each other and offer opportunities to study teacher professional knowledge more holistically (Evens et al., 2018).

Importance of Teachers' Professional Knowledge for Instructional Practice

In research on teaching effectiveness, characteristics of instructional practice have been identified in numerous studies that describe instructional quality and are beneficial for student achievement (e.g., Brophy, 2000; Seidel and Shavelson, 2007; Hattie, 2009). Moreover, there is evidence that teachers' professional knowledge about instructional practices is related to effective teaching (e.g., Baumert et al., 2010; Förtsch et al., 2016). Therefore, the acquisition of knowledge about characteristics of instructional quality should be part of teacher education. Characteristics can refer to generic or subject-specific features, and are summarized in different frameworks that cover different knowledge facets (Charalambous and Praetorius, 2018). A commonly used framework refers to the three basic dimensions of instructional quality that occur more or less across domains: classroom management, supportive climate, and cognitive activation. *Classroom management* refers both to the structure and organization of instruction and to the management of students' behavior (Klieme et al., 2001; Schlesinger and Jentsch, 2016; Praetorius et al., 2018). *Supportive climate* (also often referred to as learning support) refers to the creation of a positive learning atmosphere in the classroom. It is characterized by a caring attitude of the teacher, a positive teacher-student relationship, and other forms of support such as constructive feedback (Klieme et al., 2001; Lipowsky et al., 2009; Praetorius et al., 2018). *Cognitive activation* requires instruction that builds on students' prior knowledge and ideas, that uses challenging problems and questions to stimulate cognitive conflicts, that foster students' engagement in higher-level thinking and thus their in-depth elaboration of content, as well as students' active participation in classroom discourse (Lipowsky et al., 2009; Baumert et al., 2010; Förtsch et al., 2017). However, since many of the characteristics of cognitive activation have to be applied within subject-specific contexts, the operationalization of this basic dimension differs largely between studies due to content-specific issues of the study subject (Schlesinger and Jentsch, 2016). Furthermore, research findings revealed that "*classroom management* and *supportive climate* could be interpreted as basic conditions, which have to be established before implementing cognitively activating strategies of instruction" (Dorfner et al., 2018, *p.* 49). In conclusion, knowledge about the basic dimensions *classroom management* and *supportive climate* as well as about corresponding strategies on how to deal with these dimensions can be considered as part of teachers' PK that is important to create learning opportunities and an effective learning atmosphere in which pedagogical strategies and methods can be applied and adapted to student heterogeneity (cf. Kunter et al., 2007; König and Kramer, 2016). In contrast, *cognitive activation* is more related to knowledge of subject matter (CK and PCK) (Baumert and Kunter, 2013b). While no direct effects of CK on *cognitive activation* have been

found, there is evidence of the connectedness between CK and PCK (e.g., Krauss et al., 2008; Liepertz and Borowski, 2019). PCK, in turn, was shown to be highly predictive for instructional quality and students' achievement (Depaepe et al., 2013; Schmelzing et al., 2013; Kulgemeyer et al., 2020). An indirect effect of teachers' PCK on students' achievement mediated by *cognitive activation* was found, for example, in biology and mathematics education (Kunter et al., 2013; Förtsch et al., 2016). Recently, however, researchers have shown that teachers' PK should not be disregarded in relation to cognitive activation either, as PK was found to predict aspects of all three basic dimensions (König et al., 2021).

In addition, there are efforts to describe subject-specific characteristics such as use of technical language, dealing with student errors or conceptual instruction, the use of models and experiments in a particular subject such as mathematics or biology (Schlesinger and Jentsch, 2016; Dorfner et al., 2017; Kramer et al., 2020). Knowing about corresponding subject-specific characteristics of instructional quality is therefore related to teachers' subject-specific knowledge facets, in particular to teachers' PCK (e.g., Kunter et al., 2013).

Development of Professional Knowledge in Teacher Education

The education of pre-service teachers is based on curricula, in which the three knowledge facets CK, PK, and PCK are largely treated separately in seminars and lectures (Ball, 2000; Harr et al., 2014). On the basis of research findings from recent years, however, scientists are increasingly calling for an integrated presentation of knowledge facets, in which corresponding knowledge components are addressed together, which is assumed to improve knowledge retrievability and application in practice (Evens et al., 2018; Tröbst et al., 2019). A reason for this claim is the existing relationship between the knowledge facets. For example, CK and PK have been identified as components of PCK, but solely addressing CK and PK is not sufficient to develop PCK (e.g., König et al., 2016; Kleickmann et al., 2017; Kind and Chan, 2019). Furthermore, explicitly addressing PCK has also proven to be effective for the development of PCK (Tröbst et al., 2019). Regarding the relationship between CK and PCK, researchers emphasized that CK is considered a necessary but not sufficient condition for the development of PCK (Fischer et al., 2012; Baumert and Kunter, 2013b). In addition, other study results showed that PK is related to the instruction of specific PCK content (König et al., 2018), which also emphasizes the importance of PK in the overall knowledge development process. Furthermore, learning and retaining knowledge is considered more effective when link-making processes between new and existing ideas take place (Scott et al., 2011; Wadouh et al., 2014). Link-making can also take place in the sense that general pedagogical principles are explicitly related to subject-specific characteristics when new content is presented. In other words, link-making between PK, CK, and PCK should be given much more focus. Thus, researchers pointed out that the separate presentation of the knowledge facets might not be the most powerful way to develop

teacher professional knowledge (Evens et al., 2018; Tröbst et al., 2019).

Studies investigating the effects of an integrated presentation of the knowledge facets have already shown positive effects when using direct instructional guidance within computer-based learning environments. When creating lesson plans, integrated instructional support (content and pedagogical information were linked) was more effective than separated support (elaborate information about pedagogy and content were received separately) in terms of PCK-related justifications and the quality of PCK in lesson plans (Janssen and Lazonder, 2016). However, the authors did not include instruction on PCK as a treatment condition but only looked at PCK as an outcome variable. Harr et al. (2014) developed computer-based learning environments on mathematics and compared the effect of an integrated or segregated PCK and PK presentation on pre-service teachers' PCK and PK. "Integrated" meant that participants worked on one learning environment, in which PK and PCK aspects were treated interrelated. In contrast, in the "segregated" condition, participants worked on two learning environments, each focusing solely on either PCK or PK aspects. The results showed high effectiveness of the integrated learning environment "in increasing the application of PK aspects by pre-service teachers [...] [as well as in increasing] simultaneous application of both PCK and PK when solving a particular case from teaching practice" (Harr et al., 2014, *p.* 7). However, effects on the application of PCK aspects did not significantly differ between the integrated and the segregated condition. Furthermore, for the segregated condition, they varied the sequence of the learning environments but found no sequencing effects (Harr et al., 2014). They concluded that in teacher education, those responsible should think about how to restructure university curricula in order to allow for an integrated presentation of knowledge facets. However, since restructuring curricula that have a long tradition might be challenging and long-winded, other ways of more integration have to be found. In this vein, Harr et al. (2015) used the same methodology again but added another condition ("prompted integration") to analyze the effects of prompting pre-service teachers to integrate knowledge facets by themselves. After presenting PK and PCK separately, participants had to process prompts that asked, for example, for a connection of content topics and pedagogical principles, and were presented on additional slides to trigger mental integration. Results showed that the prompted integration was as effective as the provided integration from their first study, but at the expense of time (Harr et al., 2014, Harr et al., 2015). Nevertheless, considering feasibility for implementation in teacher education, the focus on a prompted integration might be one way to facilitate the development of pre-service teachers' professional knowledge.

While the previous studies examined only two of the three knowledge facets, Evens et al. (2018) included all three knowledge facets in their study. Situated within the subject of French as a foreign language, one question they investigated was whether a learning environment in which PCK, PK, and CK are integrated is more effective for PCK development than a learning environment in which PCK, PK, and CK are segregated. Five conditions (four experimental groups and one control group) differed in the

knowledge facets that were presented, and in the way the knowledge facets were integrated. In contrast to Harr et al. (2014, 2015), Evens et al. (2018) found no significant differences between integrated and separated instruction. In both conditions, PCK increased moderately. However, whether both groups can equally apply their knowledge to the processing of practical examples was not investigated. Furthermore, those results support previous findings on the importance of explicit instruction on PCK for the development of subject-specific knowledge. The authors also emphasized that the instruction of all three knowledge facets might then be expedient "if teacher education aims at promoting not only teachers' PCK, but also their PK and CK" (Evens et al., 2018, p. 253). Therefore, addressing different knowledge facets and connecting them can still be considered appropriate to develop pre-service teachers' professional knowledge holistically.

The Present Study

For the present study[1], we investigated the effects of both a separated and an integrated presentation of general pedagogical-psychological aspects of PK as well as aspects of the subject-specific knowledge facets CK and PCK (in terms of the subject biology). Earlier findings already indicated beneficial effects for PK and PCK when information was provided in an integrated way (e.g., Harr et al., 2014, 2015; Janssen and Lazonder, 2016) using computer-based learning environments in which the knowledge facets were differently presented (e.g., Harr et al., 2014, 2015; Janssen and Lazonder, 2016; Evens et al., 2018). To go beyond these earlier findings and to close gaps concerning the investigation of knowledge development, the present study adds value concerning two points. First, we included all three knowledge facets in our study, and we examined possible effects for PCK, as well as for CK and PK. In addition, we also included a situated measure to capture the application of professional knowledge [i.e., of applied PCK (cf. Kersting et al., 2010) or enacted PCK (Carlson et al., 2019)] by using a video-based assessment tool showing videos of biology instruction that teachers are asked to analyze (cf. Kramer et al., 2020). Second, we used a different methodological approach, which reflects common practice at universities: direct instructional guidance provided by a lecturer. A glance at university education shows that this way of supporting knowledge acquisition makes up a great deal in lectures and seminars as a common form of university courses.

Therefore, the main research question of the present study is: *Are there differences in the effectiveness of separated or an integrated instruction on the development of pre-service*

teachers' professional knowledge facets (PK, CK, and PCK) and on the application of PCK (applied PCK) in a video-based assessment tool?

We assume that integrated instruction might be more effective for PK development than separated instruction since previous findings indicated higher applicability of PK aspects when knowledge facets were acquired in an integrated way (Harr et al., 2014). However, there is also the possibility of deriving pedagogical principles from specific examples from the field of PCK, thus, enhancing PK within sequential instruction as well (König et al., 2018; Tröbst et al., 2019).

For CK, we consider the separated instruction to be more effective. CK is considered an important basis for the development of PCK. Thus, a deeper understanding of content-specific concepts and processes is the basis on which subject-specific content can adequately be prepared for students (cf. Ball et al., 2008; Krauss et al., 2008). Focusing solely on CK within instruction might help to strengthen CK without distractions.

Furthermore, since PCK contains aspects of CK and PK, we assume that the integrated instruction is more effective for the development of PCK than the sequential instruction. Our assumption is based on previous research results indicating the importance of PK and especially CK for the development of PCK (Krauss et al., 2008; Schneider and Plasman, 2011; Kleickmann et al., 2017). We assume that in the integrated instruction, two effects may be important: First, the interaction of PK and CK for the development of PCK, and the explicit instruction of PCK itself (cf. Tröbst et al., 2019).

Concerning the application of PCK, no specific assumptions can be made as there are no clear findings. On the one hand, there are studies that found no difference in the application of PCK between integrated or separated instruction (cf. Harr et al., 2014), on the other hand, there is evidence that integrated instruction is more effective for the application of PCK but only related to the integration of CK and PK (cf. Janssen and Lazonder, 2016). Therefore, the study is intended to be exploratory in this regard.

MATERIALS AND METHODS

In the original study, we used two kinds of measuring instruments: paper-pencil tests to measure the knowledge facets and the video-based assessment tool DiKoBi *Assess* (German acronym for diagnostic competences of biology teachers in biology classrooms) to measure components of diagnostic competences. Since the video-based tool presented real-life classroom situations, we took the classroom context into account that is considered to play an important role when teachers have to apply their knowledge (Kersting et al., 2010; Evens et al., 2018). Thus, measuring the professional knowledge facets PK, CK, and PCK was based on the cognitive perspective, whereas measuring diagnostic competences was based on the situated perspective (cf. Hoth et al., 2016). However, since the video-based assessment tool required that teachers apply their professional knowledge to observe and evaluate biology-specific instruction and to identify biology-specific challenges, the

---

[1]The study, presented in this article, represents a sub-analysis of the database that was collected as part of a larger study. The original study focused on the impact of integrated/separated instruction of the knowledge facets on pre-service teachers' professional knowledge and their diagnostic competences. In order to address the questions raised in the theoretical section, however, the present analysis focuses only on the effects of an integrated/separated instruction on professional knowledge. To avoid misunderstanding, the original study is referred to as "original study", whereas the present sub-analysis is referred to as "the present study" or "sub-analysis."

FIGURE 1 | Design of the study.
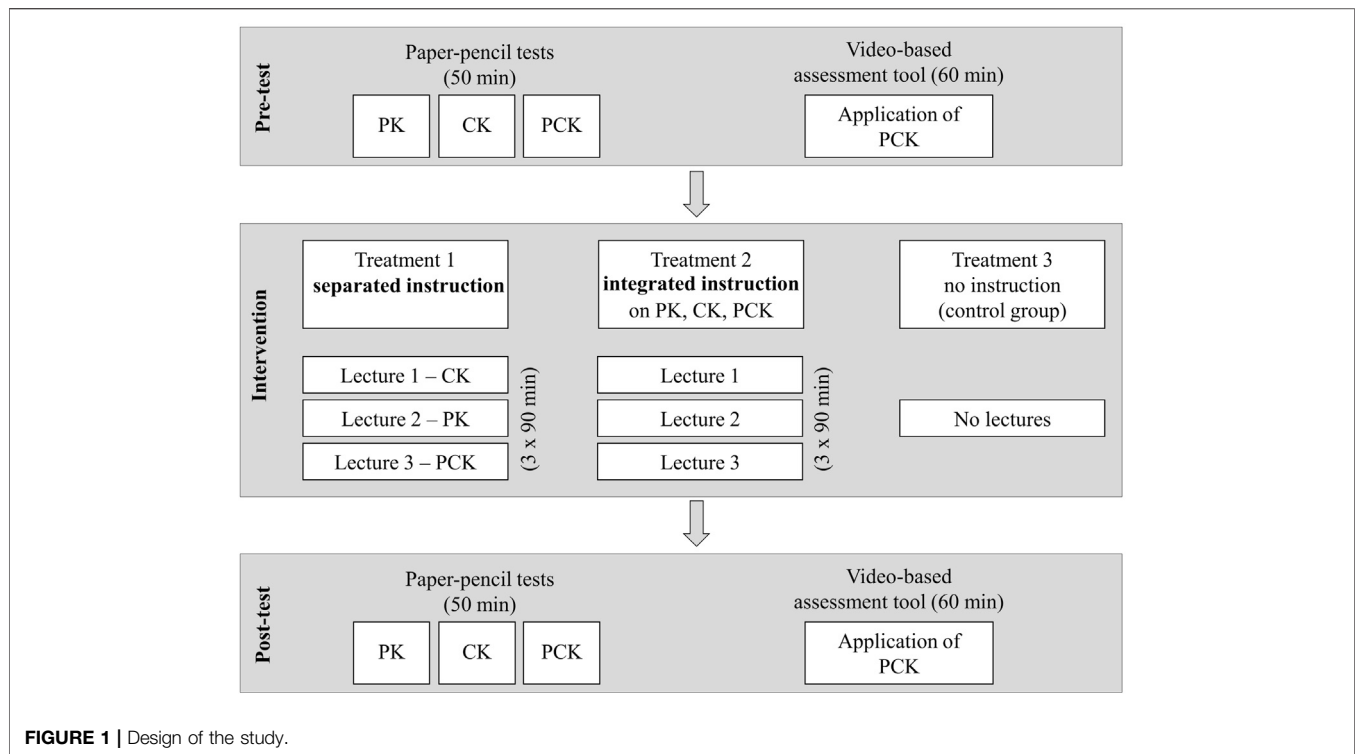
situated measurements obtained with the tool can also be considered to capture teachers' application of PCK (cf. Kersting et al., 2010; Seidel et al., 2013).

The original study focused on the topic of senses and sensory organs, which represents an important topic area within science curricula in different school types and grades (e.g., National Research Council, 2012; State Institute of School Quality and Educational Research; Munich, 2017). The topic area was exemplified for the particular topic "skin" including information on skin as a sensory organ, protective functions of the skin, and the importance of the skin for the regulation of the body temperature. The content was differentiated in such a way that aspects of the content were of practical relevance for prospective primary school teachers as well as for prospective secondary school teachers in accordance with science curricula (for more information, see Kramer et al., 2020).

Design and Sample

The study had an experimental design and was embedded in a regular seminar held once a week. The seminar is attended by pre-service biology teachers at the beginning of their teacher education. In the seminar, pre-service teachers acquire knowledge about subject-specific theories and concepts for biology instruction. The study was conducted over two weeks in May 2019, with pre-testing and post-testing during the regular seminar time. The intervention was shifted to the weekend in between. Both pre- and post-tests included three paper-pencil tests each (see **Figure 1**). Thus, each knowledge facet was measured with a separate paper-pencil test, which was the same in pre- and post-test. Additionally, the video-based

assessment tool DiKoBi *Assess* was used to measure the application of pre-service teachers' PCK in the pre- and post-test. For the intervention, pre-service teachers were randomly assigned to three different treatments. In treatment 1 and 2, a professional lecturer (first author of the article) gave three lectures on declarative and action-related aspects of PK, CK, and PCK relevant to the biological topic "skin". Each lecture took 90 min. During instruction, the different knowledge facets were addressed either in a separated or in an integrated way (see section *description of the treatments*). Pre-service teachers in treatment 3 (control group) did not receive any instruction. They only completed the pre- and the post-test. Informed consent documents stating an anonymous and voluntary participation were signed by all participants. The ethics committee of the Faculty of Psychology and Education of the LMU Munich approved the study in advance.

The sample consisted of 118 pre-service biology teachers (66.9% female; average study semester: $M = 3.02$, $SD = 1.20$; age in years: $M = 22.65$, $SD = 3.49$). 32.2% of the pre-service teachers attended the academic track of teacher education, qualifying them for future teaching at German secondary schools ("Gymnasium"); 67.8% attended programs for the non-academic track that qualifies students for vocational career. For an overview of the German school system, see Cortina and Thames (2013). **Table 1** shows how the 118 participants were distributed among the three treatments. There was no statistically significant difference in age $(F (2,114) = 2.78, p = 0.07)$, or percentage of pre-service teachers attending academic track $(F (2,115) = 0.53, p = 0.59)$. They also did not statistically differ in their knowledge at the pre-test

TABLE 1 | Overview of pre-service teachers' background characteristics within the treatments (N_{total} = 118). The mean value was given for age in years and study semester.

	Treatment		
	1 separated instruction on the knowledge facets PK, CK, PCK	**2 integrated instruction on the knowledge facets PK, CK, PCK**	**3 no instruction (control group)**
Number of participants (thereof female)	42 (29)	40 (25)	36 (25)
Age in years M (SD)	22.51 (4.18)	21.85 (2.02)	23.69 (3.73)
Study semester M (SD)	2.95 (1.19)	2.70 (1.07)	3.44 (1.25)
Percentage of pre-service teachers attending the academic track (%)	35.7	25.0	36.1

TABLE 2 | Overview of the content of the three lectures of each treatment.

	Treatment 1 separated instruction		**Treatment 2 integrated instruction**	**Treatment 3 control group**
Lecture 1 (90 min)	**CK**	– importance of the skin/basic functions – structures and functions (*epidermis* and appendages, sclera, subcutis) – the skin as a sensory organ (touch, pressure, heat, cold, pain)	**Lesson planning model – phase 1 and phase 2** – supply usage model – lesson planning model and basic dimensions of instructional quality – phases during the course of a lesson, planning instruction, subject–specific instructional quality features – phase 1: Beginning of a lesson: Classroom management, supportive climate – phase 2: Activation of prior knowledge and focus question: Level of students cognitive activities, creation of situational interest, the importance of the skin/basic functions, structures and functions of the skin (*epidermis*), cognitive activation, subject-specific theory situational interest	-
Lecture 2 (90 min)	**PK**	– supply usage model – basic dimensions of instructional quality (classroom management, supportive climate, cognitive activation in general) – lesson planning model (phases during the course of a lesson, planning instruction, teaching methods)	**Lesson planning model – phase 3** – phase 3 (part A): Elaboration and backing up the results: Students' ideas and their formative handling, structures and function of the skin (*epidermis* and appendages, sclera, subcutis), the skin as a sensory organ (touch, pressure, heat, cold, pain), conceptual change theory, classroom management, supportive climate, scientific inquiry methods, . . .	-
Lecture 3 (90 min)	**PCK**	– addressing subject-specific instructional quality features within each phase during the course of a lesson (e.g., activation of prior knowledge, use of focus questions and challenging tasks, creation of situational interest, formative handling of students' ideas, scientific inquiry methods, use of models, technical terms, linking . . .) – subject-specific theories (situational interest, conceptual change, . . .)	**Lesson planning model – phase 3, phase 4, and phase 5** – phase 3 (part B): use of models, teaching methods, technical terms – phase 4: Referring back to the focus question, linking, cognitive activation, lesson planning model, phases during the course of a lesson – phase 5: Closing the lesson	-

Note: Although the overall content was the same for both treatments, it was presented in different ways in the three lectures per group. Some of the topics in treatment 2 are listed twice because specific subtopics were addressed in one phase, while other subtopics were relevant in another phase. For example, while lecture 1 included an overview of the interdependence of classroom management, classroom context, and the teacher's personality as a subtopic of classroom management (cf. Helmke, 2017), lecture 2 included techniques of classroom management (cf. Kounin, 1970; Tarman, 2016). All lectures lasted about 90 min but varied in the number of topics addressed due to the time required to treat a topic.

[PK_{pre} (F (2,115) = 0.05, p = 0.96), CK_{pre} (F (2,115) = 0.03, p = 0.97), PCK_{pre} (F (2,115) = 0.87, p = 0.42)] or in their application of PCK at the pre-test (F (2,115) = 0.39, p = 0.61). However, a statistically significant difference was found in study semester (F (2,115) = 3.94, p = 0.02) between treatment 3 (control group) and treatment 2 (integrated instruction) as assessed by Tukey-HSD (mean difference control group-integrated instruction = 0.74, p = 0.02).

Description of the Treatments

Pre-service teachers were randomly assigned to one out of three treatments: separated instruction (treatment 1), integrated instruction (treatment 2), or no instruction/control group (treatment 3). The intervention was held on a weekend, containing three 90 min lasting lectures for treatment 1 on the weekend's Saturday, and three 90 min lasting lectures for

TABLE 3 | Example of how a task was embedded in treatment 1 and treatment 2 and how the corresponding transition was phrased.

Treatment 1 separated instruction (here: PCK)	Treatment 2 integrated instruction (here: PCK with transition to CK)
[. . .] **Task**: *Explain why it is important to consider student ideas? Collect ideas on how student ideas can be explored.* The consideration of students' ideas is important to pick up the students' level of knowledge and to motivate them according to their ideas/interests. It is also crucial to ensure that teachers and students talk about the same thing. What do students associate with the "skin"? Above all, everyday experiences are decisive for the generation of ideas. The student's idea and the academic idea of a subject are sometimes far apart. Therefore, it is important for the teacher to consider both and to structure the lessons accordingly. Didactic models that address this issue are, for example [. . .]	[. . .] The consideration of students' ideas is important to pick up the students' level of knowledge and to motivate them according to their ideas/interests. It is also crucial to ensure that teachers and students talk about the same thing. What do students associate with the "skin"? Above all, everyday experiences are decisive for the generation of ideas. The student's idea and the academic idea of a subject are sometimes far apart. Therefore, it is important for the teacher to consider both and to structure the lessons accordingly. **Task**: *Explain what becomes visible in the student's idea about the skin and what is the deficit?* To understand the student's idea about the skin, we must first familiarize ourselves with the biological content.
[*next: model of didactic reconstruction*]	[*next: structure and function of the skin*]

treatment 2 on Sunday. On both days of the weekend, the intervention was held by the same lecturer to reduce potential confounding effects. In addition, the lecturer, who was the first author of the article, prepared scripts to ensure that the overall content was kept constant in both treatments. Scripts were based on a review of the relevant literature and state-of-the-art research results of each knowledge facet with respect to the specific topic "skin". After intensive training, the lecturer strictly held the lecture according to the scripts to ensure that only the planned contents were addressed in the lectures. The distribution of the contents to the different lectures of the treatments is shown in **Table 2**.

In treatment 1, each of the lectures focused on one of the knowledge facets separately, meaning that the first lecture dealt with content knowledge (CK), the following lecture with pedagogical-psychological knowledge (PK), and the last lecture with pedagogical content knowledge (PCK). There was a 20 min break between each lecture. In each of the three lectures of treatment 2, the knowledge facets were addressed in an integrated way. The overall structure of treatment 2 followed the planning model for biology instruction (see Dorfner et al., 2019), including different phases during the course of a lesson: the beginning of a lesson (phase 1), activation of prior knowledge and focus question (phase 2), elaboration and backing up results (phase 3), referring back to the focus question, as well as a consolidation of the content/concepts being taught (phase 4), and the closing of the lesson (phase 5). In the three lectures of treatment 2, each phase of the planning model was addressed one after the other. For each phase, corresponding aspects of CK, PK, and PCK were presented (see **Table 2**). For example, the first lecture of treatment 2 (integrated instruction) opened with aspects of the basic dimensions *supportive climate* and *classroom management* that are relevant to the beginning of a lesson (referring to phase 1 of the lesson planning model). Knowing about these aspects can be considered as part of teachers' PK. The lecture continued with the content reactivation of prior knowledge (phase 2 of the lesson planning model), including teachers' dealing with student

misconceptions as part of PCK. However, in this phase, additional subject content (CK) was presented, which is necessary to identify corresponding misconceptions. Integration, therefore, followed the principle of teaching the subject content right where it is directly applicable or necessary for the understanding of a specific student misconception. The second lecture of treatment 2 (integrated instruction) opened with the elaboration phase (phase 3 of the lesson planning model). Here, the focus was on classroom management strategies to enhance time on task (PK), but also on strategies such as scientific inquiry methods relevant to implement experiments in a scientific way (PCK). In addition, appropriate knowledge necessary to understand the presented experiments was provided (CK). The third lecture of treatment 2 (integrated instruction) continued with the elaboration phase before the importance of cross-linking content was emphasized (PCK) and ways of closing a lesson were shortly presented (PK). Therefore, the third lecture referred to phase 4 and phase 5 of the lesson planning model.

All aspects of CK, PK, and PCK included in the lectures were the same for both treatments. Only the sequence of the presented knowledge facets and thus its integration varied. Therefore, the planning model mentioned above was also considered in Treatment 1 as part of the lecture on PCK.

Furthermore, we made sure to call repeatedly for pre-service teachers' attention in all lectures through short tasks that varied slightly due to an appropriate embedding in the structure of the lecture. Additionally, content connecting phrases were included in the treatments 1 and 2 to make transitions smoother. The phrases varied between the treatments but did not contain additional information on PK, CK, or PCK. **Table 3** shows an example of how a task has been embedded and how a transition has been phrased in treatment 1 and treatment 2.

Professional Knowledge Tests

PK, CK, and PCK were measured through use of paper-pencil tests. In the tests, three different types of items have been utilized (open-ended, single best answer, multiple true/false). Open-

TABLE 4 | Item examples of the CK and PCK test. Due to the different design of SBA-items, two examples are given. Examples are translated from German.

Item type	Corresponding knowledge facet	Item example
Open-ended	PCK	**Please evaluate the three-dimensional model by comparing as many different advantages and disadvantages as possible.** *Please answer the task regardless of whether you would use the model in reality.* Advantages of the model / Disadvantages of the model
Single best answer (SBA)	CK	**Read each statement and tick the statement if correct.** Statements / correct The dermis has subcutaneous fatty tissue and thus serves as protection against cold. ☐ During keratinization, the cell interior dies and is replaced by the protein carotene. ☐ […]
Single best answer (SBA)	PCK	**Please evaluate the statements below regarding the given teacher experiment.** 1 very / 2 rather / 3 neither / 4 rather not / 5 not at all a This experiment allows effective use of time. ☐ ☐ ☐ ☐ ☐ b The motivation of the students is promoted. ☐ ☐ ☐ ☐ ☐ c Students learn the essential steps of experimentation. ☐ ☐ ☐ ☐ ☐ d Practical skills for subsequent experiments are acquired. ☐ ☐ ☐ ☐ ☐ e Students can generate their own solutions. ☐ ☐ ☐ ☐ ☐

Note: Information on scales utilized in the PK test can be found in *Kunina-Habenicht et al. (2020)*. Data sets from the BilWiss project, in which the PK test utilized in this study was developed, are available on the IQB website https://www.iqb.hu-berlin.de/fdz/studies/BilWiss.

ended items required a written response in a text field; single best answer (SBA) items required the selection of a single answer from a set of possible responses consisting of multiple distractors and one correct response; multiple true/false items required the assessment of each of four given responses with respect to their correctness (Campbell, 2011). Sample items are displayed in **Table 4**.

PK was assessed by an adapted short version of a paper-pencil test utilized in the BilWiss project covering declarative knowledge of the dimension *instruction* (Kunina-Habenicht et al., 2020). According to Evens et al. (2018), PK includes at least knowledge about teaching methods and classroom management, which both were covered in the dimension *instruction* of the BilWiss project. The PK-test referred to the basic dimensions of instructional quality containing items about *classroom management*, *supportive climate*, and generic aspects of *cognitive activation* (Klieme et al., 2001; Lipowsky et al., 2009; Baumert et al., 2010), as well as items on general pedagogical issues of teaching such as teaching methods. The PK-measure included five SBA-items and ten multiple true/false items. Item scoring followed the instructions from the BilWiss project (Kunina-Habenicht

TABLE 5 | Fit statistics of the professional knowledge tests using Rasch analysis techniques and the Rasch partial credit model.

Knowledge facet	Number of items	Outfit-MNSQ	Person reliability	Item reliability
PCK$_{pre}$ (anchored)	13	<1.44	0.61	0.93
PCK$_{post}$	13	<1.15	0.63	0.95
CK$_{pre}$ (anchored)	28	<1.44	0.69	0.97
CK$_{pos}$	28	<1.43	0.69	0.97
PK$_{pre}$ (anchored)	15	<1.50	0.57	0.97
PK$_{post}$	15	<1.19	0.58	0.98

et al., 2020). SBA-items were scored with either 0 points (wrongly ticked) or 2 points (correctly ticked). Multiple true/false items were scored with either 0 points (for 0 or 1 correctly ticked answers), 1 point (for 2 or 3 correctly ticked answers), or 2 points (for 4 correctly ticked answers, thus a completely solved task).

CK and PCK were assessed by adapted versions of the professional knowledge tests used in the ProwiN project (Jüttner et al., 2013; Jüttner and Neuhaus, 2013). The CK- and PCK-test covered declarative and action-related knowledge about the topic "skin" (in accordance with the knowledge that was addressed in the intervention and also covered in the video-based assessment tool). Based on the model of Tepner et al. (2012), the PCK-test covered two important components of biology teachers' PCK: knowledge of instructional strategies (model use and use of experiments) and knowledge of students' errors. The PCK-measure included eight open-ended items and five SBA-items. Open-ended items were scored in accordance with a coding manual that was adapted from Jüttner et al. (2013) and Jüttner and Neuhaus (2012), and written on the basis of the literature in science education. A maximum of 3 points could be achieved for each open-ended item. SBA-items represented ranking items and were statements to a given experiment, which had to be evaluated on a five-point Likert scale by the pre-service teachers (see **Table 4**). Prior to the scoring process, the items were rated by 16 in-service teachers who we considered as experts in biology education. In accordance with the tendency for correct answers the in-service biology teachers gave, we divided the Likert scale into positive, neither/nor, and negative parts for scoring pre-service teachers' ratings. For example, if the mean of the experts' rating was between 1 and 2, and a pre-service teacher check-marked 1 or 2, the answer was scored with 1 point. If the pre-service teacher check-marked 3, 4, or 5, the answer was scored with 0 points (cf. Jüttner et al., 2013). The CK-measure included 13 open-ended items and 15 SBA-items. Both open-ended items and SBA-items were scored in accordance with criteria provided in a coding manual adapted from Jüttner et al. (2013). A maximum of 3 points could be achieved for each open-ended item. SBA-items were scored with either 0 points (wrongly ticked) or 1 point (correctly ticked). To ensure objective and reliable coding, ten percent of both the PCK- and CK-test were coded by two independent raters utilizing the coding manuals. A high agreement between the two raters has been shown by the results of two-way random intra-class correlations (ICC$_{absolute}$): PCK: $ICC_{absolute}$ (310,310) = 0.84, $p < 0.001$; CK: $ICC_{absolute}$ (341,341) = 0.97, $p < 0.001$ (Wirtz and Caspar, 2002).

After item scoring, each knowledge test was analyzed separately using the Rasch partial credit model (PCM), which resulted in PK, CK, and PCK Rasch person measures (person abilities) for each pre-service teacher for each test instrument (Bond and Fox, 2007; Boone et al., 2014). The Rasch model also takes the difficulty of test items into account. Person measures and item measures are calculated on the same scale using the unit "logits" as equal-interval units that allow comparisons between persons and items (Boone et al., 2014). For evaluating data fit, Outfit-MNSQ (mean-square) values, item reliability and person reliability for each test were used. According to Wright and Linacre (1994), item Outfit-MNSQ values below 1.5 indicate a productive measurement. Concerning the reliability values, high item reliability indicates that both the range of item difficulty and the sample size can be considered as appropriate to measure the items precisely. The person reliability describes the internal consistency of the measure. For example, a value of 0.50 means that the test discriminates the sample in 1 or 2 levels; higher values discriminate in more levels (Boone et al., 2014).

Item fit statistics of the PK, CK, and PCK-test showed satisfactory fit values (see **Table 5**). To compare data from the identical pre- and post-tests, pre-test items have been anchored with appropriate post-test items for each test considering Differential Item Functioning (Boone et al., 2014). Those items, which produced a measurement bias for pre- and post-test were excluded from anchoring. In the end, the PK-test included 12 anchor items, the CK-test had 17 anchor items, and the PCK-test included eight anchor items.

Video-Based Assessment Tool

In addition to the paper-pencil tests, the video-based assessment tool DiKoBi *Assess* was used in both the pre-test and the post-test. The assessment tool presented biology-specific challenges within biology instruction on the topic "skin". For the present study, we treated pre-service teachers' analysis of biology instruction as samples of applied PCK (cf. Kersting et al., 2010).

Biology-specific challenges that had to be analyzed referred to six different subject-specific dimensions that were found to be empirically effective for student achievement within science instruction (Dorfner et al., 2017). These dimensions were 1) *level of students' cognitive activities and creation of situational interest*, 2) *dealing with (specific) student ideas and errors*, 3) *use of technical language*, 4) *use of experiments*, 5) *use of models*, 6) *conceptual instruction*. The video-based assessment tool provided six videotaped classroom situations on the topic "skin as a sensory organ" for the pre-test and six videotaped classroom situations on

the topic "protective functions of the skin" for the post-test. In each situation, one of the aforementioned dimensions was addressed. Therefore, using the video-based assessment tool provides a more situated perspective on the measurement of knowledge and skills. For analysis, pre-service teachers' required content-specific pedagogical content knowledge of the characteristics of effective biology instruction that had to be applied to solve three diagnostic tasks. The diagnostic tasks were the same for all of the classroom situations that pre-service teachers watched. First, biology-specific challenges had to be identified when viewing the videotaped classroom situations and described (Task *Describe*). Second, the relevance of the identified challenges had to be justified with regard to subject-specific theories and concepts (Task *Explain*). Third, an alternative instructional strategy had to be set up (Task *Alternative Strategy*).

We consider the processes initiated by the diagnostic tasks to be similar to processes performed when pre-service teachers use their professional vision to assess classroom incidents, since professional vision "refers to the ability to notice features of a practice that are valued by a particular social group [...] and interpret instruction" (van Es and Sherin, 2008, *p.* 244). Since professional vision is considered as an indicator of integrated knowledge that teachers apply to observe and evaluate classroom instruction and to identify possible challenges (cf. Seidel et al., 2013; Seidel and Stürmer, 2014), pre-service teachers' analyses of biology instruction including biology-specific challenges are treated as a measure to capture the application of PCK (i.e., enacted PCK). From the situated perspective, the use of video-based tools is not only considered promising to measure knowledge that is activated and applied in practical situations but also to support teachers in developing knowledge and abilities that decide about teachers' effective classroom instruction (Gaudin and Chaliès, 2015; Hoth et al., 2018). Videos can capture decisive moments of classroom instruction that can serve as stimuli to elicit teachers' PCK (Kersting et al., 2010; Seidel and Stürmer, 2014; Alonzo and Kim, 2016). Thus, using video-based tools that elicit PCK may have an impact on the development of teachers' professional knowledge as well.

Pre-service teachers' answers of the three diagnostic tasks were coded in accordance with a coding manual that included indicators of subject-specific instructional quality that have been described for the six subject-specific dimensions in the science literature. Additional information about the utilized coding manual is reported in Kramer et al. (2020). Incorrect answers or those of very low quality were scored Zero (0) points. For correct answers of improved quality, 1 or 2, or with regard to Task *Explain*, 3 points could be utilized for coding. Below, examples concerning the subject-specific dimension *use of models* and corresponding scores are given.

Task *Describe*:

- "*elaborate a little more on the model*": unsystematic, superficial description (1 point)

- "*only individual parts of the model are discussed and it is not dealt with as a whole*": systematic, detailed, and complete description (2 points)

Task *Explain*:

- "*Connection between model and reality must be made*": explanatory (empty) phrase (1 point)
- "*important in the use of models is always the critical reflection of the model*": simple reference to concepts/theories (2 points)
- "*a model is presented, without explanations and critical reflection; wrong student ideas can arise as a result*": comprehensive explanation (3 points)

Task *Alternative Strategy*:

- "*briefly discuss the entire model and once again explain the function of the subcutaneous fat and that we now know where it is located*": non-specific description, rather general character (1 point)
- "*In order to avoid misconceptions among the students, address the limitations of the model (What is different about the real skin?) Comparison with a more realistic illustration*": detailed description of appropriate alternative strategy with examples (2 points)

To calculate interrater reliability, we selected all answers (covering the three diagnostic tasks for all of the six classroom situations of the pre-/post-test) from 10 randomly sampled pre-service teachers. Overall, 337 answers were coded by three independent raters. Results of a two-way random intra-class correlation ($ICC_{absolute}$) analysis suggested a high agreement between the three raters ($ICC_{absolute}$ = 0.90, F (1,520, 3,040) = 10.26, p < 0.001, N = 1,521) (Wirtz and Caspar, 2002). Discrepancies in coding were discussed by all three raters prior to the scoring of the remaining data. Complex cases continued to be discussed together during the ongoing coding process.

Afterward, coded data was used to calculate Rasch person ability measures for each respondent. Each Rasch person measure expressed the level of each pre-service teacher's ability to apply PCK in terms of describing and explaining challenging classroom situations, and proposing alternative teaching strategies. Because not exactly the same video situations were used in pre- and post-test, the person measures of both tests were treated as separate measures of applied PCK and, thus, were not anchored. Fit statistics of the Rasch model showed productive measures (application of PCK in pre-/post-test: 31^2/31 items, outfit-MNSQ ≤ 1.43/1.49; item reliability = 0.92/0.90; person reliability = 0.78/0.78).

Analyses

First, scores of all knowledge tests (pre/post: PK, CK, PCK) and scores of the assessment tool were analyzed using the Rasch PCM

---

[2]One item produced inestimable high values.

TABLE 6 | Descriptive statistics of the knowledge tests, including mean, standard deviation, minimum and maximum value.

	Treatment 1 separated instruction $N = 42$		Treatment 2 integrated instruction $N = 40$		Treatment 3 control group $N = 36$	
	Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**
PK[a]						
M	5.27	5.36	5.28	5.33	5.24	5.38
SD	0.49	0.51	0.56	0.60	0.52	0.52
Min	4.21	4.34	4.38	4.34	4.02	4.66
Max	6.49	6.48	6.49	7.28	6.21	7.28
CK[a]						
M	4.49	5.18	4.44	4.98	4.46	4.53
SD	0.68	0.81	0.78	0.76	0.81	0.67
Min	3.30	3.44	2.26	3.19	1.74	3.19
Max	5.85	6.68	5.70	6.37	5.99	5.66
PCK[a]						
M	3.83	4.37	3.62	4.35	3.79	4.23
SD	0.69	0.55	0.75	0.62	0.87	0.81
Min	2.62	3.19	1.27	2.83	1.32	2.83
Max	5.28	5.53	4.65	5.36	5.28	5.70
Applied PCK[a]						
M	3.65	3.67	3.58	3.60	3.52	3.34
SD	0.67	0.58	0.66	0.67	0.56	0.79
Min	1.25	2.26	1.90	1.67	2.28	0.56
Max	4.86	4.87	4.69	4.71	4.86	4.71

[a]Rasch person measures of variables scaled according to the PCM in Logits. Each person ability value was summed up with 5 to avoid negative person abilities and thus misunderstandings in interpretation. Note that the person measures were calibrated for each knowledge facet as well as for the application of PCK (applied PCK) separately, and therefore, they cannot be compared with each other.

(Bond and Fox, 2007) with the software Winsteps 3.81 (Linacre, 2014) to calculate person measures. The resulting equal-interval person measures were used for all following analyses. Second, descriptive results and Pearson's correlations were calculated utilizing IBM SPSS Statistics (version 26) to describe the development and correlation between the knowledge facets. SPSS has also been utilized for running mixed ANOVAs separately for each knowledge facet to reveal possible time effects and interaction effects between time and treatments and for running an ANCOVA to examine effects of the treatments on applied PCK. An ANCOVA was used because pre- and post-test differed in the topics that were addressed in the classroom situations. Still, the subject-specific dimensions to be analyzed were the same. There was a violation of normal distribution for PCK_{pre} and PK_{post} of treatment 2 (integrated instruction), for applied PCK_{pre} and applied PCK_{post} of treatment 1 (separated instruction) as well as for PK_{post} and applied PCK_{post} of the control group, as assessed by the Shapiro-Wilk test ($p < 0.05$). For PK and CK, homogeneity of the error variances, as assessed by Levene's test ($p > 0.05$), as well as homogeneity of covariances, as assessed by Box's test ($p > 0.05$) were given. However, for PCK_{post}, Levene's test was significant ($p = 0.03$). We therefore focused on Tukey-HSD post-hoc comparisons and calculated repeated measures ANOVAs for each treatment. In addition, there was homogeneity of regression slopes for applied PCK ($p = 0.49$).

As an effect size measure, we used partial η^2, applying the following benchmark values: 0.01 for small effects, 0.06 for medium effects, and 0.14 for large effects (Cohen, 1988; Richardson, 2011).

RESULTS

Descriptive Statistics and Correlations

An overview of the descriptive results of all measurements can be found in **Table 6**. For each knowledge facet, the mean values increased between pre-test and post-test. However, for PCK and CK, the increases were greater. Additionally, it is worth mentioning that for PK the control group showed the largest increase compared to treatment 1 or 2. Applied PCK remained almost at the same level for treatment 1 and 2. However, the applied PCK of the control group decreased. Pearson's correlations between the knowledge facets showed that PK, as well as CK and PCK, could mostly be measured independently. For the pre-test, there were small correlations between CK_{pre} and PCK_{pre} (r = 0.29, $p < 0.01$), as well as between PCK_{pre} and PK_{pre} (r = 0.28, $p < 0.01$). Applied PCK_{pre} showed small correlations with CK_{pre} (r = 0.26, $p < 0.01$) and PK_{pre} (r = 0.23, $p < 0.05$), and a medium correlation with PCK_{pre} (r = 0.33, $p < 0.01$). For the post-test, there was a medium correlation between CK_{post} and PCK_{post} (r = 0.40, $p < 0.01$), and a small correlation between PCK_{post} and PK_{post} (r = 0.24, $p < 0.01$). Applied PCK_{post} showed small correlations with CK_{post} (r = 0.28, $p < 0.01$) and PCK_{post} (r = 0.29, $p < 0.01$), and a medium correlation with PK_{post} (r = 0.35, $p < 0.01$).

Effects on Pedagogical-Psychological Knowledge, Content Knowledge, Pedagogical Content Knowledge, and Applied PCK

For PK, results of calculating the mixed ANOVA showed a significant effect of time (F (1, 115) = 3.94, $p = 0.05$, partial

FIGURE 2 | Comparison of pre- and post-test person measures between the different treatments, illustrated for each knowledge facet (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$). For PK and CK, mixed ANOVAs have been performed, whereas we run repeated ANOVAs for PCK. Note that the person measures were calibrated for each knowledge facet separately, and therefore, they cannot be compared with each other.

$\eta^2 = 0.03$), but no interaction effect between time and treatments ($F_{(2, 115)} = 0.23$, $p = 0.79$, partial $\eta^2 < 0.01$), meaning that pre-service teachers acquired more PK regardless of treatment affiliation. However, due to the high person measures in both the pre- and the post-test, a ceiling effect can be stated (see **Figure 2**).

Mixed ANOVAs for CK revealed that there was a statistically significant effect of time for treatment 1 (separated instruction) ($F_{(1,41)} = 28.64$, $p < 0.01$, partial $\eta^2 = 0.41$) and for treatment 2 (integrated instruction) ($F_{(1,39)} = 33.44$, $p < 0.01$, partial $\eta^2 = 0.46$). There was no significant effect of time for the control group ($F_{(1,35)} = 0.39$, $p = 0.54$, partial $\eta^2 = 0.01$). Furthermore, the interaction effect between time and treatment was significant in terms of CK-development ($F_{(2,115)} = 7.66$, $p < 0.01$, partial $\eta^2 = 0.12$). However, Tukey-HSD post-hoc tests of the mixed ANOVA revealed no significant difference between the groups. Despite

missing significance, calculation of Tukey-HSD indicated that the separated instruction might have had an advantage over the control group (mean difference separated instruction-control group = 0.34, $p = 0.07$). In contrast, this potential advantage was not apparent for the integrated instruction (mean difference integrated instruction-control group = 0.22, $p = 0.33$). However, to understand the significant interaction effect, we run another analysis. Since pre-service teachers did not statistically differ in their CK$_{pre}$, we also analyzed group differences in the post-test and calculated a one-way ANOVA for CK$_{post}$. In the post-test, CK-scores differed significantly between the treatments. According to the Tukey-HSD post-hoc tests, control group differed significantly from treatment 1 (separated instruction) (mean difference control group-separated instruction = -0.65, $p = 0.001$), and treatment 2 (integrated instruction) (mean difference control group-integrated instruction = -0.44, $p = 0.03$)

at the post-test. Overall, the results showed the effectiveness of the lectures, and might indicate a greater potential of the separated instruction in terms of CK development (see **Figure 2**).

Mixed ANOVAs for PCK could not be interpreted due to the significance of the Levene's test ($p = 0.03$), which still remained after cox-box-powertransformation (Hemmerich, 2016). Additionally, Tukey-HSD post-hoc tests showed no statistically significant difference between the treatments. However, repeated measures ANOVA for the total sample (without consideration of treatments) revealed a significant difference between PCK_{pre} and PCK_{post} ($F (1,115) = 77.76$, $p < 0.01$, partial $\eta^2 = 0.40$). Therefore, we run repeated measures ANOVA for each treatment separately. Results showed significant increases from PCK_{pre} to PCK_{post} for all three treatments (see **Figure 2**), which were much more pronounced for the treatments 1 and 2, which included instruction (treatment 1: $F (1,41) = 30.47$, $p < 0.01$, partial $\eta^2 = 0.43$; treatment 2: $F (1,39) = 44.72$, $p < 0.01$, partial $\eta^2 = 0.53$; control group: $F (1,35) = 11.38$, $p < 0.01$, partial $\eta^2 = 0.25$). The highest estimate of explained variance was reported for treatment 2, the integrated instruction (but was still high for treatment 1, the separated instruction).

Effects of treatments on the application of PCK were examined using an ANCOVA. Results showed that the covariate applied PCK_{pre} was significantly related to applied PCK_{post} ($F (1,114) = 55.12$, $p < 0.001$, partial $\eta^2 = 0.33$). There was no significant effect of treatment on PCK_{post} after controlling for the effects of the covariate ($F (2,114) = 2.46$, $p = 0.09$, partial $\eta^2 = 0.04$), meaning that pre-service biology teachers in all treatments did not significantly differ in their applied PCK after the intervention.

DISCUSSION

In the present study, we compared the effects of separated or integrated instruction on PK, CK, and PCK on the development of pre-service teachers' professional knowledge and the application of knowledge in terms of PCK. Of critical significance was that all three knowledge facets were addressed both as part of the instruction and as outcome variables, whereas previous studies were often limited to specific knowledge facets (cf. Harr et al., 2014; Janssen and Lazonder, 2016). Regarding this point, our study provides insights into how teachers' professional knowledge can be fostered. Furthermore, we investigated the integration of the knowledge facets in a very practical way within regularly scheduled lectures. Therefore, the study also provides practical value for realizations of teacher education programs. Although the use of computer-based learning environments offers advantages in terms of controllability and standardizability, we decided to investigate the integration of the knowledge facets in lectures in order to expand methodological approaches within the investigation of knowledge instruction. In addition, study results could be used directly to adapt instruction due to the affiliation to our courses. Therefore, we addressed how curricular content should be presented in courses and lectures of science teacher education. Despite the evidence of greater effectiveness of integrated

instruction (Harr et al., 2015; Janssen and Lazonder, 2016), knowledge facets are largely taught in separate courses in teacher education. Since curricular restructuring is not feasible without enormous effort, the purpose of this study was to generate more evidence on the extent to which restructuring in lectures might be effective. For this purpose, three treatments (separated instruction, integrated instruction, and control group receiving no instruction) were compared on participants' PK, CK, and PCK development as well as on their application of PCK.

The following main findings of the statistical analysis can be noted: All knowledge facets (PK, CK, PCK) increased from pre-to post-test. The largest increases were shown with respect to the development of CK and PCK. However, not all increases can be attributed to the intervention. Referring to PK, a small time effect but no interaction effect between time and treatments could be found. Instead, we noticed a ceiling effect, indicating that the utilized PK-test has been too easy and did not differentiate the sample enough (Linacre, 2014). Furthermore, the descriptive results showed the largest PK-increases in the control group. This might indicate a testing effect (Shadish et al., 2002), possibly boosted by the test's insufficient complexity. On the other hand, the increases in the control group might also be due to the use of the video-based tool. Although the videos showed subject-specific instruction, general pedagogical aspects are recognizable to a certain extent and general pedagogical aspects may be derived from the subject-specific implementation (cf. Tröbst et al., 2019). For example, one classroom situation dealt with specific student misconceptions on the subject of skin. It is conceivable that general strategies for favorable handling of student errors, which were asked in the knowledge test, might potentially be derived from the specific situation shown in the videos. An indication of the potential relationship between PK and the application of knowledge in the video-based tool may also be seen in the increase in correlation between PK and applied PCK from pre to post-test. In the future, we should raise test difficulty through an extension of the PK-test by including other scales of the original BilWiss project (Kunina-Habenicht et al., 2020). Thus, at this stage, we cannot make any firm conclusions about differences in the effectiveness of separated or integrated instruction on the development of PK.

Referring to the development of CK, both addressing knowledge facets separately or integrated in lectures were effective. There were large interaction effects between time and treatments of both the separated and the integrated instruction compared to the control group that showed no significant difference between CK_{pre} and CK_{post}. Evidence that either separated or integrated instruction is more effective can only be identified descriptively but should be interpreted with caution. The descriptive results indicated that the separated instruction was slightly more effective for CK development. This seems plausible because before teachers can teach a topic, they first need knowledge about the subject matter and they have to understand the underlying core concepts before actually planning instruction (Ball et al., 2008; Kleickmann et al., 2017). Furthermore, pre-service teachers can be considered as novices for whom less integration of

knowledge facets is characteristic, in contrast to the knowledge of experienced in-service teachers, which is stronger integrated and encapsulated (Krauss et al., 2008). Therefore, due to little subject-specific knowledge about a specific topic and due to the still less integrated knowledge structures, the separated instruction might be the more effective one in terms of CK development. However, the present study could not provide significant evidence for this, which might be due to the small sample size of the treatments, which reduced statistical power. Future studies that take the different levels of prior knowledge of a higher sample size of pre-service teachers into account could provide more differentiated insights on the circumstances of a beneficial instruction on CK.

Referring to the development of PCK, it was striking that a large time effect was found not only in the two treatment groups but also in the control group, who did not receive any kind of instruction during the intervention. This effect may be due to the processing of the biology-specific classroom situations provided in the video-based assessment tool. The use of the tool seems to have had an impact on the knowledge facets captured as defined by the cognitive perspective. To explain this observation, we want to refer to the Refined Consensus Model mentioned in the theoretical section to define the PCK-constructs used in our study more precisely. Thus, the paper-pencil tests utilized in the pre- and post-tests measured pre-service teachers' subject-specific knowledge (*personal* PCK) that reflected their person-specific reservoir of declarative and action-related knowledge as well as individual teaching and learning experiences (Carlson et al., 2019). When elaborating on the PCK aspects of the intervention, we focused on topic-specific literature within the field and state-of-the-art research results. Thus, this realm of PCK represented *collective* PCK that interacted with pre-service teachers' *personal* PCK and was therefore assumed to change it. Additionally, we assume that pre-service teachers had to rely on their *personal* PCK when working in the video-based assessment tool during the diagnostic process while simultaneously drawing on *enacted* PCK that is generated in the moment of action (Alonzo et al., 2019). Even if pre-service teachers were not actually in action themselves, they engaged in the practice of science teaching in terms of reflecting on biology instruction, and thus, utilized *enacted* PCK. It is assumed that through reflection, *enacted* PCK can be transformed to *personal* PCK, and thus, the experiences can become part of future knowledge (Alonzo et al., 2019). Since we used videos of real-life classroom instruction, the videos and their reflection might have elicited knowledge that could then be accessed in the post-test. This could have been, for example, knowledge about the use of three-dimensional models triggered by the situated context in the video. The videos thus functioned as an additional prompt for the retrieval of *personal* PCK (Kersting et al., 2010; Seidel and Stürmer, 2014; Alonzo and Kim, 2016). Therefore, the results of our sub-analysis also emphasized the relevance of considering both a cognitive and a situated perspective on teacher professional knowledge. Considering the results, there is the possibility that an interaction effect between time and

treatments was overshadowed by the effect of the video-based assessment tool. However, effect sizes suggest higher effects for the two treatments that received instruction. Though the value of explained variance was greater for the integrated instruction, a statistically significant advantage of the integrated instruction could not be found compared to other studies (Harr et al., 2014, Harr et al., 2015). Still, the higher value of explained variance for integrated instruction might indicate that PCK development might benefit slightly more from the interrelated instruction on PK and CK, as well as from the explicit instruction on PCK itself (Krauss et al., 2008; Schneider and Plasman, 2011; Kleickmann et al., 2017; Tröbst et al., 2019).

Since pre-service teachers' application of PCK did not change after the intervention, in which declarative and action-related knowledge was presented by a lecturer, more practice-oriented training forms may be more beneficial, since simply acquiring professional knowledge (either separated or integrated) might not be sufficient for enacting the knowledge to diagnose subject-specific instruction (cf. Kron et al., 2021). Supporting the application of PCK directly in the video-based assessment tool might be one way to improve the development of integrated knowledge that is applicable to instructional situations. Motivational conditions would then also have to be controlled to counteract a possible decrease in motivation in the post-test (as may have occurred in the control group in the present study).

Despite these insights, there are some limitations to the present study. First, some assumptions relevant for applying mixed ANOVAs were violated. However, the violation of normal distribution for PCK_{pre} and PK_{post} of treatment 2 (integrated instruction), applied PCK_{pre} and applied PCK_{post} of treatment 1 (separated instruction) as well as for PK_{post} and applied PCK_{post} of the control group, can be explained by the fact that PCK_{pre} as well as the application of PCK measured with the video-based assessment tool was rather difficult for the pre-service teachers. Thus, due to the small group sizes of treatments, the probability that the normal distribution is violated increased. Results of PCK_{post} tests were normal distributed for all treatments. PK_{post}, on the other hand, was too easy, which is why there was no normal distribution for two treatments there either. Since pre-service teachers in our study were at the beginning of their studies, the application of knowledge is still challenging and, thus, the amount of integrated, enactable knowledge is rather low (Kind and Chan, 2019). A second limitation concerns the knowledge tests used. Since we used the same paper-pencil test for pre- and post-test, a test effect might have occurred. Additionally, knowledge increases of the knowledge facets, even when not significant, might be due to pre-service teachers' regular courses that took place during the study but were not part of it. Regarding the ceiling effect resulting from the utilized PK-test, we cannot make any statements on the effects of the treatments on PK, which is why the discussion refers primarily to CK and PCK findings. Therefore, an extended test version should be piloted and used in future studies. Accordingly, the current study design would have to be applied again in a similar way using an extended PK test

version in order to conduct and discuss the targeted, holistic investigation of all three knowledge facets. This is especially important to evaluate whether integrated instruction might also be favorable in terms of PK development. If this is the case, university courses would need to pursue a stronger interlocking of the more practice-oriented facets PK and PCK (cf. König et al., 2018). At the same time, expanding the study could result in a larger sample size of the individual treatments. The increase in sample size might then provide more clarity about the partly descriptive trends in PCK and CK. Conducting the study at other universities would also be conceivable. However, it is a great challenge to embed the lecture-based study in exactly the same form at other universities since conception and implementation were strongly oriented toward the instructional practice of biology education at the LMU Munich. However, this also represents a strength, as research and practice have been combined.

Furthermore, we could only measure specific aspects of a pre-service teacher's knowledge facet referring to articulable knowledge "related to the teaching and learning of specific science topics" (Alonzo et al., 2019, p. 273). The paper-pencil tests did not allow us to capture more dynamic forms of knowledge that are used in practice, quasi "in action" (Alonzo and Kim, 2016), and thus, the measurement of professional knowledge with paper-pencil tests might lack sufficiency (Liepertz and Borowski, 2019). However, measurement data from the video-based tool provided more information about the application of PCK as dynamic, integrated form of knowledge (cf. Kersting et al., 2010; Seidel et al., 2013), but effects on the application of CK or PK could not be examined due to the focus of the videotaped classroom situations that lied on subject-specific challenges of biology instruction. With reference to the RCM, it further remains open which particular realm of PCK was addressed and to what extent integrated instruction is equally effective for different realms. Consequently, future studies would do well not only to differentiate into knowledge facets, but also to consider potential realms of a knowledge facet (as in the case of the PCK realms) that is relevant to interpretation and eventual consequences.

Another relevant limitation is the absence of a manipulation check of the treatments. Thus, we cannot ensure that pre-service teachers in the treatments that received instruction had processed the presented knowledge. However, in the lectures, attempts were made to control this point. Since the sample size of the treatments was not too big, and the lecture hall, although rather small, provided enough seats to distribute the participants evenly, the lecturer could keep an eye on them. Furthermore, the lecturer tried to make sure that no participant was distracted by secondary activities (mobile phone, conversations) but instead was stimulated by tasks to ensure active and constructive participation at least temporarily, which is assumed to foster learning (Chi and Wylie, 2014). Nevertheless, the cognitive presence of the participants has not been instrumentally controlled.

Nonetheless, our study contributes to the exploration of integrated approaches to promote knowledge acquisition.

Therefore, we finally want to point out some implications resulting from our findings. Overall, direct instructional guidance provided by a lecturer can be considered as effective way of knowledge acquisition. However, the strategy should not limit oneself to one way of knowledge instruction. The choice of a specific way of instruction should depend, among other things, on the level of prior knowledge of the pre-service teachers. Further studies that investigate separated and integrated approaches with regard to developmental trajectories of pre-service teachers with different levels of prior knowledge should be initiated in the future. As the present analysis has shown, direct instructional guidance as provided through lectures is an efficient way to foster pre-service biology teachers' CK that builds the crucial fundament for a profound development of PCK. Addressing CK separately might therefore be a way to ensure that pre-service teachers develop a sufficient level of CK that impacts PCK development in the longer term, and thus, instructional quality as well (Baumert et al., 2010). Within the instruction of PCK, it might be practicable to refer to subject-specific content and core concepts as well as to general pedagogical methods that can be transferred to the specific subject to be taught. This can be done in a way that Harr et al. (2015) called prompted integration describing the use of reflective questions to promote knowledge integration. There is the necessity to bring the knowledge facets together in order to increase their applicability in complex instructional situations (Ball, 2000; Harr et al., 2015; Tröbst et al., 2019). These situations may occur during classroom instruction as well as when instruction is planned or reflected. In order to address this practical context already in teacher education at university, situated approaches on knowledge acquisition and application are increasingly in demand. Video-based tools such as the assessment tool that offers real-life classroom situations can therefore be seen as instructional elements with practical relevance that complement teacher education. There is great potential to combine approaches to integrated instruction of knowledge with the use of video-based tools, not only in research contexts (cf. Harr et al., 2014; Janssen and Lazonder, 2016), but also directly in teacher education courses.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the DFG-funded project and the final data preparation are not yet completed. Requests to access the datasets should be directed to the authors of this article.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of the Faculty of Psychology and Education of the LMU Munich. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FUNDING

REFERENCES

Alonzo, A. C., and Kim, J. (2016). Declarative and Dynamic Pedagogical Content Knowledge as Elicited through Two Video-Based Interview Methods. *J. Res. Sci. Teach.* 53 (8), 1259–1286. doi:10.1002/tea.21271

Alonzo, A. C., and Kim, J. (2018). Affordances of Video-Based Professional Development for Supporting Physics Teachers' Judgments about Evidence of Student Thinking. *Teach. Teach. Educ.* 76, 283–297. doi:10.1016/j.tate.2017.12.008

Alonzo, A. C., Berry, A., and Nilsson, P. (2019). "Unpacking the Complexity of Science Teachers' PCK in Action: Enacted and Personal PCK," in *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*. Editors A. Humel, R. Cooper, and A. Borowski (GatewayEast, Singapore: Spinger), 273–288. doi:10.1007/978-981-13-5898-2_12

Ball, D. L., Thames, M. H., and Phelps, G. (2008). Content Knowledge for Teaching: What Makes it Special?. *J. Teach. Educ.* 59 (5), 389–407. doi:10.1177/0022487108324554

Ball, D. L. (2000). Bridging Practices: Intertwining Content and Pedagogy in Teaching and Learning to Teach. *J. Teach. Educ.* 51 (3), 241–247. doi:10.1177/0022487100051003013

Baumert, J., and Kunter, M. (2013a). The COACTIV Model of Teachers' Professional Competence. *Cogn. activation Math. classroom Prof. competence Teach.*, 25–48. doi:10.1007/978-1-4614-5149-5_2

Baumert, J., and Kunter, M. (2013b). The Effect of Content Knowledge and Pedagogical Content Knowledge on Instructional Quality and Student Achievement. *Cogn. activation Math. classroom Prof. competence Teach.*, 175–205. doi:10.1007/978-1-4614-5149-5_9

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *Am. Educ. Res. J.* 47 (1), 133–180. doi:10.3102/0002831209345157

Blömeke, S., Gustafsson, J.-E., and Shavelson, R. J. (2015). Beyond Dichotomies. *Z. für Psychol.* 223 (1), 3–13. doi:10.1027/2151-2604/a000194

Bond, T. G., and Fox, C. M. F. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2nd Edn.. Mahwah, NJ: Lawrence Erlbaum Associates.

Boone, W. J., Staver, J. R., and Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Berlin, Germany: Springer. doi:10.1007/978-94-007-6857-4

Brophy, J. (2000). *Teaching*. Educational Practices Series. Available at: http://www.ibe.unesco.org/fileadmin/user_upload/archive/Publications/educationalpracticesseriespdf/prac01e.pdf (Accessed March 30, 2021).

Brunner, M., Anders, Y., Hachfeld, A., and Krauss, S. (2013). "The diagnostic skills of mathematics teachers," in *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*. Editors M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand (New York: Springer) 229–248. doi:10.1007/978-1-4614-5149-5_11

Campbell, D. E. (2011). How to Write Good Multiple-Choice Questions. *J. Paediatrics Child Health* 47 (6), 322–325. doi:10.1111/j.1440-1754.2011.02115.x

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., et al. (2019). The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. *Repositioning Pedagogical Content Knowl. Teachers' Knowl. Teach. Sci.*, 77–94. doi:10.1007/978-981-13-5898-2_2

Cauet, E., Liepertz, S., Borowski, A., and Fischer, H. E. (2015). Does it Matter what We Measure? Domain-specific Professional Knowledge of Physics Teachers. *Swiss J. Educ. Res.* 37 (3), 463–480. doi:10.24452/sjer.37.3.4963

Charalambous, C. Y., and Praetorius, A.-K. (2018). Studying Mathematics Instruction through Different Lenses: Setting the Ground for Understanding Instructional Quality More Comprehensively. *ZDM Math. Educ.* 50 (3), 355–366. doi:10.1007/s11858-018-0914-8

Chi, M. T. H., and Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol.* 49 (4), 219–243. doi:10.1080/00461520.2014.965823

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Cortina, K. S., and Thames, M. H. (2013). "Teacher education in Germany," in *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project*. Editors M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand (New York: Springer) 49–62. doi:10.1007/978-1-4614-5149-5_3

Depaepe, F., Verschaffel, L., and Kelchtermans, G. (2013). Pedagogical Content Knowledge: A Systematic Review of the Way in Which the Concept Has Pervaded Mathematics Educational Research. *Teach. Teach. Educ.* 34, 12–25. doi:10.1016/j.tate.2013.03.001

Dorfner, T., Förtsch, C., and Neuhaus, B. J. (2017). Die Methodische und Inhaltliche Ausrichtung Quantitativer Videostudien zur Unterrichtsqualität im mathematisch-naturwissenschaftlichen Unterricht. *ZfDN* 23 (1), 261–285. doi:10.1007/s40573-017-0058-3

Dorfner, T., Förtsch, C., and Neuhaus, B. J. (2018). Effects of Three Basic Dimensions of Instructional Quality on Students' Situational Interest in Sixth-Grade Biology Instruction. *Learn. Instruction* 56, 42–53. doi:10.1016/j.learninstruc.2018.03.001

Dorfner, T., Förtsch, C., Spangler, M., and Neuhaus, B. J. (2019). Wie plane ich eine konzeptorientierte Biologiestunde? Ein Planungsmodell für den Biologieunterricht. - Das Schalenmodell - [How Do I plan concept-based biology instruction? A lesson planning model for biology instruction]. *Der Mathematische Und Naturwissenschaftliche Unterricht (MNU)* 4, 300–306.

Evens, M., Elen, J., Larmuseau, C., and Depaepe, F. (2018). Promoting the Development of Teacher Professional Knowledge: Integrating Content and Pedagogy in Teacher Education. *Teach. Teach. Educ.* 75, 244–258. doi:10.1016/j.tate.2018.07.001

Fischer, H. E., Borowski, A., and Tepner, O. (2012). Professional Knowledge of Science Teachers. *Second Int. Handbook Sci. Educ.*, 435–448. doi:10.1007/978-1-4020-9041-7_30

Förtsch, C., Werner, S., von Kotzebue, L., and Neuhaus, B. J. (2016). Effects of Biology Teachers' Professional Knowledge and Cognitive Activation on Students' Achievement. *Int. J. Sci. Educ.* 38 (17), 2642–2666. doi:10.1080/09500693.2016.1257170

Förtsch, C., Werner, S., Dorfner, T., von Kotzebue, L., and Neuhaus, B. J. (2017). Effects of Cognitive Activation in Biology Lessons on Students' Situational Interest and Achievement. *Res. Sci. Educ.* 47 (3), 559–578. doi:10.1007/s11165-016-9517-y

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M., Girwidz, R., Obersteiner, A., et al. (2018). Systematizing Professional Knowledge of Medical Doctors and Teachers: Development of an Interdisciplinary Framework in the Context of Diagnostic Competences. *Educ. Sci.* 8 (4), 207. doi:10.3390/educsci8040207

Gaudin, C., and Chaliès, S. (2015). Video viewing in teacher education and professional development: a literature review. *Educ. Res. Rev.* 16, 41–67. doi:10.1016/j.edurev.2015.06.001

Gess-Newsome, J. (2015). "A model of professional knowledge and skill including PCK: Results of the thinking from the PCK Summit," in *Re-examining pedagogical content knowledge in science education*. Editors A. Berry, P. Friedrichsen, and J. Loughran (New York: Routledge), 28–42.

Harr, N., Eichler, A., and Renkl, A. (2014). Integrating Pedagogical Content Knowledge and Pedagogical/psychological Knowledge in Mathematics. *Front. Psychol.* 5, 1–10. doi:10.3389/fpsyg.2014.00924

Harr, N., Eichler, A., and Renkl, A. (2015). Integrated Learning: Ways of Fostering the Applicability of Teachers' Pedagogical and Psychological Knowledge. *Front. Psychol.* 6, 738–816. doi:10.3389/fpsyg.2015.00738

Hattie, J. (2009). *Visible Learning: A Synthesis of Meta-Analyses Relating to Achievement*. London, United Kingdom: Routledge.

Heitzmann, N., Seidel, T., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., et al. (2019). Facilitating Diagnostic Competences in Simulations in Higher Education A Framework and a Research Agenda. *Flr* 7, 1–24. doi:10.14786/flr.v7i4.384

Helmke, A. (2017). Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts [Instructional quality and teacher professionalism. Diagnosis, evaluation and improvement of teaching] (7. Auflage). Schule weiterentwickeln, Unterricht verbessern Orientierungsband. Klett-Kallmeyer.

Hemmerich, W. A. (2016). StatistikGuru: Box-Cox Powertransformation Berechnen. Available at: https://statistikguru.de/rechner/box-cox.html (Accessed December 17, 2020).

Hill, H. C., Ball, D. L., and Schilling, S. G. (2008). Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-specific Knowledge of Students. *J. Res. Math. Educ.* 39 (4), 372–400.

Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., and Blömeke, S. (2016). Diagnostic Competence of Primary School Mathematics Teachers during Classroom Situations. *ZDM* 48 (1-2), 41–53. doi:10.1007/s11858-016-0759-y

Hoth, J., Kaiser, G., Döhrmann, M., König, J., and Blömeke, S. (2018). A Situated Approach to Assess Teachers' Professional Competencies Using Classroom Videos. *Mathematics Teach. Engaging Representations Pract. A Dynamically Evolving Field*, 23–45. doi:10.1007/978-3-319-70594-1_3

Janssen, N., and Lazonder, A. W. (2016). Supporting Pre-service Teachers in Designing Technology-Infused Lesson Plans. *J. Comput. Assist. Learn.* 32 (5), 456–467. doi:10.1111/jcal.12146

Jüttner, M., and Neuhaus, B. J. (2012). Development of Items for a Pedagogical Content Knowledge Test Based on Empirical Analysis of Pupils' Errors. *Int. J. Sci. Educ.* 34 (7), 1125–1143. doi:10.1080/09500693.2011.606511

Jüttner, M., and Neuhaus, B. J. (2013). Validation of a Paper-And-Pencil Test Instrument Measuring Biology Teachers' Pedagogical Content Knowledge by Using Think-Aloud Interviews. *J. Educ. Train. Stud.* 1 (2), 113–125. doi:10.11114/jets.v1i2.126

Jüttner, M., Boone, W., Park, S., and Neuhaus, B. J. (2013). Development and Use of a Test Instrument to Measure Biology Teachers' Content Knowledge (CK) and Pedagogical Content Knowledge (PCK). *Educ. Asse Eval. Acc.* 25 (1), 45–67. doi:10.1007/s11092-013-9157-y

Keller, M. M., Neumann, K., and Fischer, H. E. (2017). The Impact of Physics Teachers' Pedagogical Content Knowledge and Motivation on Students' Achievement and Interest. *J. Res. Sci. Teach.* 54 (5), 586–614. doi:10.1002/tea.21378

Kersting, N. B., Givvin, K. B., Sotelo, F. L., and Stigler, J. W. (2010). Teachers' Analyses of Classroom Video Predict Student Learning of Mathematics: Further Explorations of a Novel Measure of Teacher Knowledge. *J. Teach. Educ.* 61 (1-2), 172–181. doi:10.1177/0022487109347875

Kind, V., and Chan, K. K. H. (2019). Resolving the Amalgam: Connecting Pedagogical Content Knowledge, Content Knowledge and Pedagogical Knowledge. *Int. J. Sci. Educ.* 41 (7), 964–978. doi:10.1080/09500693.2019.1584931

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006). Why Minimal Guidance during Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educ. Psychol.* 41 (2), 75–86. doi:10.1207/s15326985ep4102_1

Kleickmann, T., Tröbst, S., Heinze, A., Bernholt, A., Rink, R., and Kunter, M. (2017). Teacher Knowledge Experiment: Conditions of the Development of Pedagogical Content Knowledge. *Methodol. Educ. Meas. Competence Assess. Educ.*, 111–129. doi:10.1007/978-3-319-50030-0_8

Klieme, E., Schümer, G., and Knoll, S. (2001). "Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung [Mathematics Instruction in Secondary Education: Task Culture and Instructional Processes]," in *Bundesministerium für Bildung und Forschung. TIMMS - Impulse für Schule und Unterricht [TIMSS – impetus for school and teaching* (Bundesministerium für Bildung und Forschung (BMBF)), 43–57.

König, J., and Kramer, C. (2016). Teacher Professional Knowledge and Classroom Management: on the Relation of General Pedagogical Knowledge (GPK) and Classroom Management Expertise (CME). *ZDM* 48 (1-2), 139–151. doi:10.1007/s11858-015-0705-4

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., and Kaiser, G. (2014). Is Teachers' General Pedagogical Knowledge a Premise for Noticing and Interpreting Classroom Situations? A Video-Based Assessment Approach. *Teach. Teach. Educ.* 38, 76–88. doi:10.1016/j.tate.2013.11.004

König, J., Lammerding, S., Nold, G., Rohde, A., Strauß, S., and Tachtsoglou, S. (2016). Teachers' Professional Knowledge for Teaching English as a Foreign Language. *J. Teach. Educ.* 67 (4), 320–337. doi:10.1177/0022487116644956

König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M., et al. (2018). Pädagogisches Wissen versus Fachdidaktisches Wissen?. *Z. Erziehungswiss* 21 (3), 1–38. doi:10.1007/s11618-017-0765-z

König, J., Blömeke, S., Jentsch, A., Schlesinger, L., née Nehls, C. F., Musekamp, F., et al. (2021). The Links between Pedagogical Competence, Instructional Quality, and Mathematics Achievement in the Lower Secondary Classroom. *Educ. Stud. Math.* 102 (1), 1. doi:10.1007/s10649-020-10021-0

Kounin, J. S. (1970). *Discipline and Group Management in Classrooms*. New York, NY: Holt, Reinhardt & Winston.

Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., and Neuhaus, B. J. (2020). Measuring Biology Teachers' Professional Vision: Development and Validation of a Video-Based Assessment Tool. *Cogent Educ.* 7 (1). doi:10.1080/2331186x.2020.1823155

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (2008). Pedagogical Content Knowledge and Content Knowledge of Secondary Mathematics Teachers. *J. Educ. Psychol.* 100 (3), 716–725. doi:10.1037/0022-0663.100.3.716

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., et al. (2013). "Mathematics teachers' domain-specific professional knowledge: conceptualization and test construction in COACTIV," in *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project.* Editors M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand (New York: Springer), 147–174. doi:10.1007/978-1-4614-5149-5_8

Kron, S., Sommerhoff, D., Achtner, M., and Ufer, S. (2021). Selecting Mathematical Tasks for Assessing Student's Understanding: Pre-service Teachers' Sensitivity to and Adaptive Use of Diagnostic Task Potential in Simulated Diagnostic One-To-One Interviews. *Front. Educ.* 6, 738. doi:10.3389/feduc.2021.604568

Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., et al. (2020). Professional Knowledge Affects Action-related Skills: The Development of Preservice Physics Teachers' Explaining Skills during a Field Experience. *J. Res. Sci. Teach.* 4 (3), 1105. doi:10.1002/tea.21632

Kunina-Habenicht, O., Maurer, C., Wolf, K., Holzberger, D., Schmidt, M., Dicke, T., et al. (2020). Der BilWiss-2.0-Test: Ein revidierter Test zur Erfassung des bildungswissenschaftlichen Wissens von (angehenden) Lehrkräften [[The BilWiss-2.0 Test: A Revised Instrument for the Assessment of Teachers' Educational Knowledge]. *Diagnostica* 66 (2), 80–92. doi:10.1026/0012-1924/a000238

Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., et al. (2007). "Linking aspects of teacher competence to their instruction: results from the COACTIV project," in *Studies on the educational quality of schools: The final report on the DFG Priority Programme*. Editor M. Prenzel (Münster: Waxmann), 32–52.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., and Hachfeld, A. (2013). Professional Competence of Teachers: Effects on Instructional Quality and Student Development. *J. Educ. Psychol.* 105 (3), 805–820. doi:10.1037/a0032583

Liepertz, S., and Borowski, A. (2019). Testing the Consensus Model: Relationships Among Physics Teachers' Professional Knowledge, Interconnectedness of Content Structure and Student Achievement. *Int. J. Sci. Educ.* 41 (7), 890–910. doi:10.1080/09500693.2018.1478165

Linacre, J. M. (2014). A User's Guide to Winsteps® Ministeps Rasch-Model Computer Programs: Program Manual 3.81.0. Available at: http://www.winsteps.com/manuals.htm (Accessed June 06, 2020).

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., and Reusser, K. (2009). Quality of Geometry Instruction and its Short-Term Impact on

Students' Understanding of the Pythagorean Theorem. *Learn. Instruction* 19 (6), 527–537. doi:10.1016/j.learninstruc.2008.11.001

National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C., WA: National Academies Press.

Park, S., and Oliver, J. S. (2008). Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Res. Sci. Educ.* 38 (3), 261–284. doi:10.1007/s11165-007-9049-6

Praetorius, A.-K., Klieme, E., Herbert, B., and Pinger, P. (2018). Generic Dimensions of Teaching Quality: the German Framework of Three Basic Dimensions. *ZDM Math. Educ.* 50 (3), 407–426. doi:10.1007/s11858-018-0918-4

Renkl, A., Mandl, H., and Gruber, H. (1996). Inert Knowledge: Analyses and Remedies. *Educ. Psychol.* 31 (2), 115–121. doi:10.1207/s15326985ep3102_3

Richardson, J. T. E. (2011). Eta Squared and Partial Eta Squared as Measures of Effect Size in Educational Research. *Educ. Res. Rev.* 6 (2), 135–147. doi:10.1016/j.edurev.2010.12.001

Schlesinger, L., and Jentsch, A. (2016). Theoretical and Methodological Challenges in Measuring Instructional Quality in Mathematics Education Using Classroom Observations. *ZDM* 48 (1-2), 29–40. doi:10.1007/s11858-016-0765-0

Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A., and Neuhaus, B. J. (2013). Development, Evaluation, and Validation of a Paper-And-Pencil Test for Measuring Two Components of Biology Teachers' Pedagogical Content Knowledge Concerning the "Cardiovascular System". *Int. J. Sci. Math. Educ.* 11 (6), 1369–1390. doi:10.1007/s10763-012-9384-6

Schneider, R. M., and Plasman, K. (2011). Science Teacher Learning Progressions. *Rev. Educ. Res.* 81 (4), 530–565. doi:10.3102/0034654311423382

Scott, P., Mortimer, E., and Ametller, J. (2011). Pedagogical Link-making: a Fundamental Aspect of Teaching and Learning Scientific Conceptual Knowledge. *Stud. Sci. Educ.* 47 (1), 3–36. doi:10.1080/03057267.2011.549619

Seidel, T., and Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Rev. Educ. Res.* 77 (4), 454–499. doi:10.3102/0034654307310317

Seidel, T., and Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *Am. Educ. Res. J.* 51 (4), 739–771. doi:10.3102/0002831214531321

Seidel, T., Blomberg, G., and Renkl, A. (2013). Instructional Strategies for Using Video in Teacher Education. *Teach. Teach. Educ.* 34, 56–65. doi:10.1016/j.tate.2013.03.004

Sekretariat der Ständigen Konferenz der Kultusminister der Ländern der Bundesrepublik Deutschland (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Beschluss vom 16.12.2004 [Educational standards in biology for the intermediate school certificate]. München, Germany: Wolters Kluwer.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educ. Res.* 15, 4–14. doi:10.3102/0013189x015002004

Shulman, L. (1987). Knowledge and Teaching:Foundations of the New Reform. *Havard Educ. Rev.* 57, 1–23. doi:10.17763/haer.57.1.j463w79r56455411

State Institute of School QualityEducational Research Munich(2017). LehrplanPLUS Mittelschule: Natur und Technik 5. Available at: https://www.lehrplanplus.bayern.de/fachlehrplan/mittelschule/5/nt/regelklasse#77436 (Accessed December 21, 2020).

Tarman, B. (2016). Discipline or Classroom Management. *J. Learn. Teach. Digital Age* 1 (2), 37–44. doi:10.46303/ressat.01.01.4

Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., et al. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften [Item Development Model for Assessing Professional Knowledge of Science Teachers]. *Z. für Didaktik der Naturwissenschaften* 18, 7–28. Available at: http://archiv.ipn.uni-kiel.de/zfdn/pdf/18_Tepner.pdf (Accessed April 27, 2021).

Tröbst, S., Kleickmann, T., Depaepe, F., Heinze, A., and Kunter, M. (2019). Effects of Instruction on Pedagogical Content Knowledge about Fractions in Sixth-Grade Mathematics on Content Knowledge and Pedagogical Knowledge. *Unterrichtswiss* 47 (1), 79–97. doi:10.1007/s42010-019-00041-y

van Es, E. A., and Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teach. Teacher Edu.* 24 (2), 244–276. doi:10.1016/j.tate.2006.11.005

Voss, T., and Kunter, M. (2013). Teachers' General Pedagogical/Psychological Knowledge. *Cogn. activation Math. classroom Prof. competence Teach.*, 207–227. doi:10.1007/978-1-4614-5149-5_10

Voss, T., Kunter, M., and Baumert, J. (2011). Assessing Teacher Candidates' General Pedagogical/psychological Knowledge: Test Construction and Validation. *J. Educ. Psychol.* 103 (4), 952–969. doi:10.1037/a0025125

Wadouh, J., Liu, N., Sandmann, A., and Neuhaus, B. J. (2014). The Effect of Knowledge Linking Levels in Biology Lessons upon Students' Knowledge Structure. *Int. J. Sci. Math. Educ.* 12, 25–47. doi:10.1007/s10763-012-9390-8

Wirtz, M. A., and Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. [Interrateragreement and interraterreliability]*. Göttingen, Germany: Hogrefe.

Wright, B. D., and Linacre, J. M. (1994). Reasonable Mean-Square Fit Values. Available at: https://www.rasch.org/rmt/rmt83b.htm (Accessed June 06, 2020).