

Article

Can Pre-Service Biology Teachers' Professional Knowledge and Diagnostic Activities Be Fostered by Self-Directed Knowledge Acquisition via Texts?

Maria Kramer ^{*} , Christian Förtsch  and Birgit J. Neuhaus 

Biology Education, Department I, Faculty of Biology, LMU Munich, 80797 Munich, Germany; christian.foertsch@bio.lmu.de (C.F.); didaktik.biologie@lrz.uni-muenchen.de (B.J.N.)

* Correspondence: maria.kramer@biologie.uni-muenchen.de

Abstract: In a diagnostic context of reasoning about instructional quality, scientific reasoning skills can be described as diagnostic activities, which require professional knowledge. Different approaches to enhance pre-service teachers' professional knowledge (PCK, CK, PK), as well as diagnostic activities exist. However, results about their effectiveness are still inconsistent. We systematically investigated the effectiveness of self-directed knowledge acquisition via texts on PCK, CK, PK, and diagnostic activities of 81 pre-service biology teachers following an experimental design. Paper-pencil tests, measuring PCK, CK, and PK, and the video-based assessment tool DiKoBi Assess, measuring diagnostic activities in the context of diagnosing instructional quality, were used pre and post an intervention. Intervention included four treatments on self-directed knowledge acquisition via texts on (1) PCK, (2) CK, (3) PK, (4) combination PCK/CK/PK. Treatment (5) served as control. Mixed ANOVAs showed large time effects for PCK and CK, but no interaction effect concerning knowledge facets between time and treatment for any of the treatments. Time effects might be due to pre-service teachers' scientific reasoning on biology instruction that activated knowledge. An ANCOVA showed no significant effect of treatment on diagnostic activities either. We conclude that scientific reasoning about instructional quality is more effective for knowledge acquisition than text-work.

Keywords: professional knowledge; scientific reasoning skills; diagnostic activities; knowledge acquisition; pedagogical content knowledge; video-based assessment; biology education



Citation: Kramer, M.; Förtsch, C.; Neuhaus, B.J. Can Pre-Service Biology Teachers' Professional Knowledge and Diagnostic Activities Be Fostered by Self-Directed Knowledge Acquisition via Texts?. *Educ. Sci.* **2021**, *11*, 244. <https://doi.org/10.3390/educsci11050244>

Academic Editor:
Stephen A. Butterfield

Received: 26 March 2021
Accepted: 13 May 2021
Published: 19 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Scientific reasoning, as a component of scientific inquiry, encompasses reasoning, and problem-solving processes that count as crucial for coping with science-related issues of everyday-life [1,2]. Therefore, Krell et al. [3] underlined the importance of developing scientific reasoning competencies during teacher education. Krell et al. defined scientific reasoning competencies as a complex construct comprising three knowledge types for problem-solving (*knowing that, knowing how, knowing why*, cf. [4]) that are applied in cognitive processes (e.g., encoding, strategy development, cf. [2]). For this application, scientific reasoning skills, such as formulating questions, testing hypotheses, planning, and performing investigations, analyzing information systematically, and drawing reasonable conclusions from specific observations are required [2,5,6]. With regard to a rather broad understanding of scientific reasoning competencies, they can also be considered important for teachers to monitor and improve instructional quality in the science classroom. In such a context of instructional diagnosis, a teacher explicitly and systematically compares different characteristics of instruction in a data-based manner in order to be able to make appropriate instructional decisions [7]. Scientific reasoning competencies in such a context of diagnosis can be considered similar to conceptualizations of diagnostic competences [8,9]. For diagnostic competences, different components such as professional knowledge and diagnostic activities have been distinguished. Knowledge has been conceptualized in terms

of the content-related facets pedagogical content knowledge (PCK), content knowledge (CK), and psychological-pedagogical knowledge (PK), as well as with regard to types of knowledge (*knowing that, knowing how, knowing when and why*) [10]. Since knowledge is applied in diagnostic contexts, and since within those contexts, the diagnostic focus can vary, all content-related knowledge facets may be of importance. Therefore, *knowing that, knowing how, knowing when and why* is not restricted to CK, but can also be distinguished for PCK and PK, which can be seen as an extension of the definition given in the scientific reasoning competencies approach. The conceptualization of diagnostic activities in which knowledge is applied in order to solve specific problems can be seen as equivalent to scientific reasoning skills [8].

Whereas many studies in recent years have examined the structure of professional knowledge and developed instruments to measure different knowledge facets (e.g., COACTIV: *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* [11]; ProwiN: *Professional Knowledge of Teachers in Science* [12]), there is still a lack on how to effectively promote the content-related knowledge facets PCK, CK, and PK, and whether there are differences regarding the effectiveness of procedures such as text work, lectures, or practical training. Additionally, the effects of fostering professional knowledge on the execution of scientific reasoning skills are not well studied either. Within a biology context, the present study addresses this issue by investigating whether pre-service biology teachers' professional knowledge can be supported by self-directed knowledge acquisition via texts that represents one common working method in higher education programs and, thus, has high practical relevance for pre-service teachers. The study further investigates whether this text-based support affects the pre-service teachers' scientific reasoning skills about subject-specific instruction expressed as diagnostic activities.

The following theoretical section starts from the professional competence of teachers, from which corresponding conceptualizations of knowledge and skills are derived that are relevant for scientific reasoning in the context of diagnosis.

1.1. Conceptualizing Teachers' Professional Competence in Terms of Diagnosing

Teachers' professional competence has been studied from a cognitive perspective focusing dominantly on teachers' knowledge facets (e.g., [11,13,14]) and from a situated perspective including the context in which instructional decisions have to be made (e.g., [15–17]). The competence as a continuum model combines both perspectives and defines professional competence as a continuum with different components spanning from a teacher's dispositions (e.g., professional knowledge, beliefs) that underlie situation-specific skills, which in turn inform the teacher's actual instruction [18]. Dispositions are defined as "underlying characteristic of a person" [19] (p. 97) that can be regarded as cognitive in terms of professional knowledge and as affect-motivation in terms of teachers' belief, interest, or motivation [18,20]. Conceptualizations of situation-specific skills refer to teachers' adequate coping with teaching situations and allow an action-oriented assessment of variables that take situated learning approaches into account [20,21]. In the context of reasoning, such skills can be operationalized as scientific reasoning skills [22].

Furthermore, teachers' instruction in the classroom is considered to be decisive for teaching effectiveness. Effective teaching can be described by generic and subject-specific dimensions of instructional quality, which influence students' outcomes [23–25]. As part of their professional competence, teachers should know about these dimensions and included instructional quality features in order to reason about them, to diagnose instructional processes, make appropriate decisions, and adapt teaching (cf., [15,26]). Generic instructional quality features can be described by three basic dimensions of instructional quality: *classroom management*, *supportive climate*, and *cognitive activation*. The dimensions *classroom management* (includes strategies and activities such as rules, routines, or monitoring to organize the classroom and ensure an effective use of time) and *supportive climate* (includes, amongst others, a teacher's sensitivity to learners, patient teacher actions, establishment of a positive learning climate, and appropriate feedback) are understood as generally applica-

ble across domains. In contrast, the content-dependent basic dimension *cognitive activation* (includes instructional practices which stimulate students to higher cognitive engagement to foster conceptual understanding) is to be differentiated more subject-specifically and, thus, is considered closer to subject-specific dimensions [27,28]. Relevant aspects of *cognitive activation* that have been described are conceptual instruction and an appropriate level of students' cognitive activities [24]. However, teachers of a specific subject, such as biology, do not only have to reason about the basic dimensions of instructional quality, but above all they must be able to describe and implement instructional key features of the specific subject, since it is those subject-specific instructional quality features that are considered necessary for high-quality biology instruction (cf. [27,29]). Diagnosing biology-specific instructional quality features (e.g., teachers' formative handling with specific student ideas, the thoughtful use of content-specific technical language, an elaborate use of models in order to solve scientific questions or the application of scientific inquiry strategies when planning and conducting experiments) is therefore of great importance [23,25,30].

Within the situation-specific processes of a subject such as biology, teachers evaluate data (e.g., from monitoring scientific inquiry steps, the elicitation of student thinking, the use of three-dimensional models) to inform their pedagogical reasoning and decision-making [31]. Since such processes are part of the systematic and continuous generation and evaluation of knowledge about students and (subject-specific) instructional dimensions, they can be summarized as diagnosing, which counts as an important component of teachers' professional competence [32,33]. Taking the data-based process into account, diagnosing can be considered as a type of scientific reasoning including several epistemic activities teachers can make use of [34]. In the context of diagnosis, these activities have also been described as diagnostic activities [8]. Therefore, teacher education should not only foster teachers' professional knowledge regarding instructional quality features but also enable them to "apply their knowledge in diagnostic activities according to professional standards to collect and interpret data in order to make decisions of high quality" [8] (p. 9). Thus, diagnostic activities represent scientific reasoning skills used to specify situation-specific skills in the context of diagnosis.

Depending on the context in which knowledge has to be applied, different content-related knowledge facets and types of knowledge may be critical [10]. In addition, different diagnostic activities have been described, but not all of them are considered relevant for diagnosing specific situations [8], since in some situations it might be sufficient to generate and evaluate evidence, whereas in other situations the generation of hypothesis or the creation or redesign of artifacts may be more important for knowledge generation. In the following, different approaches to conceptualize teacher professional knowledge and skills are described.

1.1.1. Conceptualizing Teachers' Professional Knowledge

Effective biology teaching requires different knowledge types. Förtsch et al. [10] distinguished knowledge related to facts, terms, and principles (mostly referred to as declarative knowledge or *knowing that*), and action-related knowledge (*knowing how*, *knowing when and why*). *Knowing that* means, for example, that a teacher can correctly list the advantages and disadvantages of a specific model. *Knowing how* refers to knowledge about actions, procedures, or manipulations, and is applied, for example, when a teacher deals with students' ideas. *Knowing when and why* relates to knowledge about when and why to apply particular procedures to achieve particular goals, for example, knowing when and why students' errors within a certain topic are dealt with [10,35].

In addition to the division into different types of knowledge, content-related knowledge facets can be classified. Based on Shulman's classification [36,37], most models focused on PCK, CK, and PK that build the core of the construct [11]. Knowledge can be described for all three knowledge facets in terms of declarative and action-related knowledge [10]. The subject-independent facet PK contains knowledge that counts as necessary for classroom management, classroom assessment, and organization to facilitate

an effective learning atmosphere in which pedagogical strategies can be applied [26,38,39]. Broader conceptualizations, as used in the BilWiss project, include the PK-dimensions instruction (which is further divided into the sub-dimensions generic instructional quality features and teaching strategies and methods), learning and development, diagnostics and evaluation, educational theory, school as an educational institution, and teaching as a profession [40]. In general, the conditions a science teacher establishes in the classroom are assumed to provide the basis on which PCK and CK can be used [39]. Since PCK and CK count as subject-specific, they are most important for science education. CK describes the knowledge of subject matter, discipline-specific methods for generating knowledge, and the conceptual understanding of specific topics, which researchers emphasized as a necessary but insufficient precondition for the development of PCK [11,36,41]. With regard to a specific subject matter, PCK includes subject-specific knowledge about corresponding (mis)conceptions of particular students, knowledge about subject-specific structures of instruction, and corresponding teaching strategies, and was shown to be highly predictive for instructional quality and students' achievement [15,21,42–44]. Accordingly, PCK is related to the implementation of subject-specific instructional quality features and is thus considered particularly relevant for subject-specific instruction [42,44]. However, researchers emphasized that it is not only this stage of knowledge that forms PCK but also the knowledge that is closely related to the actual practice and, thus, is more dynamic [43,45,46]. Science teaching includes taking students' prior learning into account, facilitating linkages between concepts, or choosing and utilizing instructional strategies that best suit particular teaching moments. Such tendencies underpin teachers' pedagogical reasoning, which is the heart of teaching, and are regarded as components of PCK rather than equivalent to PCK (see [47,48]).

For measuring professional knowledge, researchers mostly used paper-pencil tests with validated test items [11,49]. However, to measure context-dependent, practice-oriented knowledge, other approaches than paper-pencil assessments are needed that take the situated character of knowledge application into account [50]. The use of video analyses represents such an approach [51].

1.1.2. Conceptualizing Situation-Specific Skills for Reasoning about Instruction

Researchers assume that teachers' professional knowledge underlies their situation-specific skills that teachers use to systematically solve specific situations in the classroom or to inform subject-specific instruction [18]. When solving (problematic) situations, teachers engage in reasoning processes in order to make decisions. From a scientific stance, "scientific reasoning encompasses the reasoning and problem-solving skills involved in generating, testing and revising hypotheses or theories, and in the case of fully developed skills, reflecting on the process of knowledge acquisition and knowledge change that results from such inquiry activities" [2] (p. 61). In this definition, it becomes evident that scientific reasoning comprises specific processes that aim at generating knowledge [2,34]. The mentioned processes are in line with Nowak et al. [5], who identified three main processes that are central to scientific reasoning: (1) asking questions and formulating hypotheses, (2) planning and performing an investigation, (3) analyzing data and reflecting on the investigation. For effective engagement in any of these reasoning processes, scientific reasoning skills including reasoning from evidence are required [22,52]. Several skills have been described either with regard to the three main processes (cf. [5]) or with regard to epistemic activities that have been found relevant for generating knowledge in different domains [34]. Overall, eight epistemic activities have been described: (1) *problem identification*, (2) *questioning*, (3) *hypothesis generation*, (4) *construction and redesign of artefacts*, (5) *evidence generation*, (6) *evidence evaluation*, (7) *drawing conclusions*, and (8) *communicating and scrutinizing*. Such scientific reasoning skills are considered vital not only for teachers' classroom instruction [53] but for every human's understanding of the world and for the development of responsible citizenship [2,54]. Accordingly, it is important that science

teachers master such skills: Firstly, to promote them among their students and, secondly, to apply them in the context of reasoning processes about science or biology instruction.

Within the broader understanding of scientific reasoning, applying scientific reasoning on instruction can be seen as an evidence-based process of systematically collecting data, generating and evaluating evidence, and drawing inferences in order to produce a diagnosis and make instructional decisions (cf. [8,31,55]). Therein, evidence is not only the product of an experimental investigation but can also consist of statements describing observations (cf. [52]). Overall, scientific evidence-based reasoning in the context of diagnosing enables teachers to diagnose their students and also instructional features that are important in terms of instructional quality, and thus, for student learning (cf. [11,42]). In teacher education, pre-service teachers should, therefore, also develop knowledge and skills to enact scientific reasoning in different contexts, such as diagnosing, for which specific tools are needed [3].

In the context of diagnosing, scientific reasoning skills have been operationalized as diagnostic activities. Diagnostic activities describe those activities that teachers execute for data-evaluation within situation-specific diagnostic contexts and that are more clearly observable than solely cognitive processes that underlie diagnosing [8,56]. Diagnostic activities have been mentioned in several studies (e.g., [57,58]), but an explicit definition of different activities is mostly missing. A more differentiated approach was recently made by Heitzmann et al. [8], who translated the eight epistemic activities introduced by Fischer et al. [34] into eight diagnostic activities relevant for the goal-oriented process of diagnosing. In the following, these diagnostic activities are illustrated with examples from teaching:

- *identifying problems* (e.g., a teacher recognizes a noteworthy incident in classroom instruction driven by prior knowledge, cf. [59]);
- *questioning* (e.g., a teacher asks for reasons of the identified problematic incident);
- *generating hypothesis* (e.g., a teacher makes an assumption about the underlying problem of the teaching situation);
- *constructing artifacts* (e.g., a teacher generates tests/tasks to be used for (further) data collection);
- *generating evidence* (e.g., a teacher or observer uses the test or task or systematically observes and describes the situation, for example, with regard to relevant student or teacher behavior);
- *evaluating evidence* (e.g., a teacher interprets data and evaluates the extent to which it supports a demanded standard);
- *drawing conclusions* (e.g., a teacher derives (behavioral) consequences from the evaluation of multiple data sources);
- *communicating the process and results* (e.g., a teacher shares findings and feedback can be given; afterward, further measures can be taken or alternative instructional strategies can be implemented).

The eight diagnostic activities can be understood as a reservoir of activities teachers can use for diagnosis. Which diagnostic activities are appropriate may differ with regard to the discipline or the diagnostic focus [8,60]. In studies that investigated diagnostic activities in the context of teacher education, the diagnostic activities *generating hypothesis*, *generating evidence*, *evaluating evidence*, and *drawing conclusions* were considered to be particularly relevant [55,60,61]. Furthermore, it is noteworthy that some diagnostic activities show similarities with conceptualizations of situation-specific cognitive skills, such as perceiving, interpreting, and decision-making (PID model, [51]) or professional vision [62,63]. Both conceptualizations have been used in the context of video analysis and the diagnosis of classroom instruction. From the perspective of the PID-model, teachers' abilities to *perceive* particular events in instructional settings, to *interpret* the events, and to *make decisions* either as anticipating answers to student ideas or proposing alternative teaching strategies have been identified as crucial in terms of professional competence. From the perspective of professional vision, the skills of *noticing* (paying attention to noteworthy

events) and *reasoning* (describing noteworthy events, explaining by linking pedagogical concepts and principles to observed events, and predicting possible consequences as specification of teachers' decision making) are highlighted [63]. Therefore, researchers using diagnostic activities in the context of video analysis have operationalized *generating evidence*, *evaluating evidence*, and *drawing conclusions* as reasonable diagnostic activities for assessment designs [61].

Even when scientific reasoning skills, and more explicit the epistemic purpose underlying a diagnostic activity, are transferable across disciplines [64], to a certain extent, the application of scientific reasoning skills is discipline- and context-specific, as well as knowledge-dependent (cf. [65,66]). Therefore, the adequate execution of diagnostic activities in the context of diagnosing biology instruction may rely on subject-specific facets of professional knowledge such as teachers' PCK (cf. [9,29]).

1.2. Fostering Professional Knowledge and (Scientific) Reasoning Skills

Due to the complex interaction and interdependence of professional knowledge and scientific reasoning skills, such as diagnostic activities, pre-service teachers need varying opportunities to develop knowledge and apply diagnostic activities during their teacher education. The common division into three content-related knowledge facets is also reflected in the university organization, in which the knowledge facets are taught in separate courses and lectures [67], while different knowledge types are addressed more or less explicitly across all courses. Standard working methods in higher education include text-based procedures requiring pre-service teacher to acquire knowledge self-directed, lecture-based procedures in which specific information is presented by a lecturer, mixed forms of text- and lecture-based instruction, and situated approaches to learning that represent scenarios from real-world demands, for example, in video vignettes (cf. [68,69]). Depending on the context, on learners' prerequisites, on structure and content of materials, and specific learning goals, the effects of instructional support on knowledge acquisition may vary.

Barth et al. [70] compared the effects of self-directed knowledge acquisition and direct instruction of knowledge about classroom disruptions that represent an aspect of PK on three cognitive outcomes: (declarative) knowledge on classroom management, noticing critical incidents in the classroom, and knowledge-based reasoning. Results showed that direct instruction that was conducted by a university teacher and included a systematic introduction to the relevant content led to higher gains in knowledge on classroom management (PK) and improved the "ability to apply this knowledge in a simulated teaching situation (the video) through knowledge-based reasoning" [70] (p. 8). However, noticing (that corresponds to the diagnostic activity *problem identification*) was not affected. In addition, the self-directed acquisition of knowledge did not result in any significant effects. Kleickmann et al. [71] investigated conditions that are necessary for developing PCK in mathematics education. Experimentally manipulated treatments received instructions by an experienced lecturer on different combinations and sequencing of declarative and action-related PCK, CK, or PK over two days. Additionally, important content was repeated, and the participating pre-service teachers received handouts and had to carry out different tasks, such as answering short questions or writing assignments to recapitulate the major contents of the instruction. Regarding direct instruction, their results showed that "explicitly addressing the knowledge of students, learning and teaching in concrete content domains, whether with or without antecedent CK instruction, appeared to be the most effective pathway" [71] (p. 126). A reanalysis of the data also underlined that instruction on PCK has small effects on CK or PK development as well [67]. The authors of the study attributed this to comparisons and reflections stimulated by the test questions that might have prompted particular aspects of CK and PK. Furthermore, Smit et al. [72] investigated the relationships between PCK, CK, and scientific inquiry attitudes. They found gains for both declarative PCK and CK measured with test items within a training program on scientific inquiry. To ensure all participants had a similar level of knowledge, a

teacher educator gave a theoretical PCK input on scientific inquiry. Furthermore, inquiry-related videos were discussed. Results showed a major relationship between PCK and scientific inquiry attitudes. In addition, the input proved to be effective for PCK and scientific inquiry attitudes. The ensuing intervention consisted of a peer-coaching on lesson planning focusing on scientific inquiry skills. However, lesson planning was not found to affect professional knowledge. Possible effects on scientific inquiry skills have not been investigated.

Besides studies focusing on the three knowledge facets PCK, CK, or PK, training interventions that aim to improve situation-specific skills exist. Positive effects of university courses were shown, in which noticing and knowledge-based reasoning were fostered by using videos [63,73,74]. In most of these courses, pre-service teachers discussed and reflected teaching performances shown in video clips, which potentially addressed the diagnostic activities *problem identification*, *generating* and *evaluating evidence*, and *drawing conclusions*. To our knowledge, effects of instructional support via texts explicitly on activities relevant within diagnostic contexts have not been investigated yet.

However, a first approach to differentiate between measurements of cognitive dispositions in terms of knowledge facets and situated skills was recently made by Gess-Newsome et al. [45]. As part of a three-year professional development training, participants studied curriculum materials, discussed issues of effective pedagogy, and deepened CK to promote different facets of knowledge and skills. Investigated facets of teachers' knowledge and skills were declarative CK, general PK, two separate PCK-constructs (PCK-PK and PCK-CK), as well as inquiry-oriented teacher practice. Even though general PK was conceptualized as a cognitive knowledge facet in their initial conception, the authors finally recognized that by using an observation protocol to assess PK in video-recorded classroom sessions, they actually measured a skill instead of declarative PK. Furthermore, the assessment of the PCK-constructs was situated and can be considered as an approach to elicit skills as enacted form of PCK (cf. [48]). These PCK-related skills reflected the application of reasoning skills in terms of diagnostic activities on a meta-level. Skills referred to the abilities to describe a lesson (i.e., *generating evidence*), to explain rationales for instruction (i.e., *evaluating evidence*), and to make instructional decisions (i.e., *drawing conclusions*). In addition, Gess-Newsome et al. [45] investigated the development of teachers' inquiry-based instruction and, thus, included a second measure related to reasoning skills directly used in action. As the result of the three-year professional development training, the authors found an increase in all investigated facets of teachers' knowledge and skills, indicating the effectiveness of integrating multiple pathways of teachers' professional learning.

Despite the number of studies carried out in the field of professional competence training, only a few studies can be found that explicitly investigated a certain type of knowledge acquisition (self-directed via texts, lecture or instruction, video-club) and its impact on the knowledge facets PCK, CK, and PK, as well as on teachers' skills such as diagnostic activities in a systematic way. Previous studies have rarely focused on the investigation and support of all knowledge facets equally, nor have skills such as diagnostic activities been investigated with regard to subject-specific instructional quality features. Moreover, in some studies, the conceptualization of investigated variables lacks preciseness or has not been clearly considered.

Assessment situations should include the explicit measurement of all knowledge facets and types or at least situate measures to specific facets and types for a more fine-grained differentiation and analysis of professional knowledge in order to examine knowledge development in relation to methodological approaches and training measures (cf. [10,75]). This is important for understanding the nature of the individual knowledge facets and skills, and for clarifying whether and what type of intervention is effective. It also provides important information for practical implementation.

1.3. Motivation of the Study and Research Questions

Teacher education at universities continues to provide a great deal of knowledge acquisition about text-based instruction, although it is unclear to what extent this is effective. The question arises whether this kind of learning setting is best for pre-service teachers with little classroom experience. Therefore, our first goal was to investigate the effects of knowledge acquisition via texts and how self-directed knowledge acquisition via texts affects pre-service teachers' professional knowledge facets PCK, CK, and PK. Our second goal was to analyze if knowledge acquisition also affects the application of scientific reasoning skills within the context of diagnosing subject-specific instructional quality. Since it is assumed that the use of scientific reasoning skills, such as diagnostic activities, relies to some degree on an individual's knowledge (cf. [65]), there might not only be an effect of self-directed knowledge acquisition via texts on the professional knowledge facets but on diagnostic activities as well (cf. [15,48]). Thus, the present study addresses the following research questions (RQ):

- RQ1a: Is the self-directed knowledge acquisition via texts effective to foster pre-service biology teachers' PCK, CK, and PK?
- RQ1b: Are there different effects of the intervention on pre-service biology teachers' PCK, CK, and PK?
- RQ2: Is the self-directed knowledge acquisition via texts effective to foster the execution of diagnostic activities?

2. Materials and Methods

2.1. Design and Sample

The present study was embedded in pre-service biology teacher's university studies within a regular seminar held once a week. It took a total of three seminar dates in autumn 2018. The seminar dealt with basic theories and concepts for teaching biology and is attended by pre-service biology teachers at the beginning of their teacher education. Using the video-based tool *DiKoBi Assess* (German acronym for diagnostic competences of biology teachers in biology classrooms) was compulsory for all seminar attendees. However, consent to the use of data for analysis was voluntary. All participants signed informed consent documents stating an anonymous and voluntary participation.

The experimental design of the study contained a pre-test (day 1), a post-test (day 3) and featured five different treatments (intervention on day 2). Pre- and post-data were collected in two steps each (see Figure 1). First, pre-service biology teachers completed three paper-pencil tests to measure their PCK, CK, and PK, which were the same in pre and post-test. Second, we used the video-based assessment tool *DiKoBi Assess* to measure pre-service teachers' diagnostic activities pre (*DiKoBi I Assess*) and post (*DiKoBi II Assess*). The assessment tool provides videotaped classroom situations that have to be diagnosed with regard to different subject-specific dimensions (see Section 2.3.2 Video-Based Assessment Tool *DiKoBi Assess*). The diagnostic tasks were the same for *DiKoBi I* and *II*; both versions differed only in the content of the classroom situations shown. Both pre- and post-measurements took 120 min each. The intervention lasted 90 min and consisted of five different treatments, in which information on either (1) PCK, (2) CK, (3) PK, (4) a combination of these three knowledge facets, or (5) none of these knowledge facets (control group) was acquired in a self-directed way. The intervention covered declarative and action-related knowledge relevant for teaching the topic "skin". The same topic was also addressed in the videotaped classroom situations of the assessment tool [76].

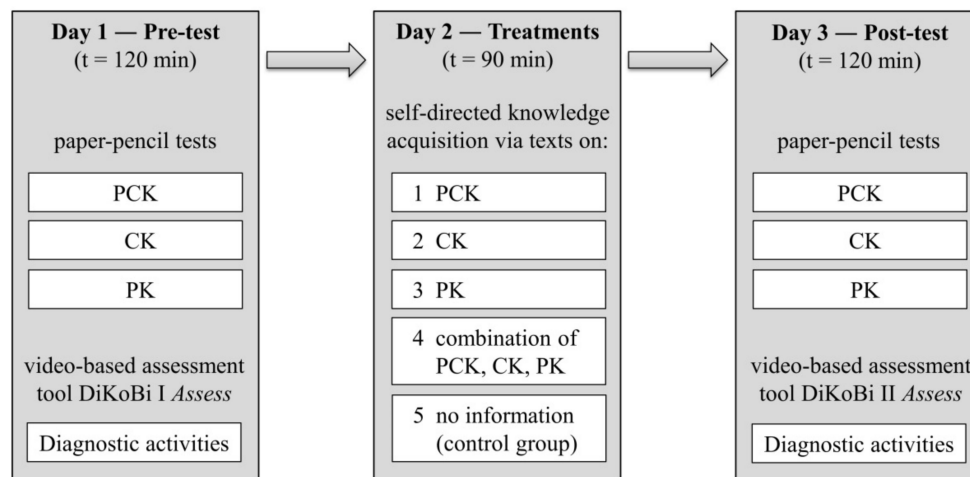


Figure 1. Design of the study.

The sample consisted of 81 pre-service biology teachers (75.3% female; average study semester: $M = 3.9$, $SD = 1.3$; age in years: $M = 23.6$, $SD = 3.9$). Overall, 48.1% of the pre-service teachers attended the academic track of teacher education, qualifying them for future teaching at German secondary schools (“Gymnasium”); 51.9% attended programs for the non-academic track that prepares students for a vocational career (for an overview of the German school system see [77]).

Pre-service teachers were randomly assigned to five treatments (see Table 1). There was no statistically significant difference in age ($F(4, 76) = 0.77$, $p = 0.55$), study semester ($F(4, 76) = 0.55$, $p = 0.70$), or percentage of pre-service teachers attending academic track ($F(4, 76) = 0.62$, $p = 0.65$). They also did not statistically differ in their pre PCK ($F(4, 76) = 0.81$, $p = 0.52$), pre CK ($F(4, 76) = 0.21$, $p = 0.93$), pre PK ($F(4, 76) = 0.33$, $p = 0.86$), or pre diagnostic activities ($F(4, 76) = 1.23$, $p = 0.31$).

Table 1. Descriptive statistics of the five treatments ($N_{\text{total}} = 81$). Mean value (M) and standard deviation (SD) are given for age in years, study semester, and the results of the pre-tests.

	Treatment				
	1 PCK	2 CK	3 PK	4 Combination of PCK, CK, PK	5 No Information (Control Group)
Number of participants (thereof female)	15 (13)	15 (11)	16 (15)	17 (11)	18 (11)
Age in Years M (SD)	24.4 (5.0)	23.6 (4.4)	24.6 (5.6)	22.9 (2.0)	22.7 (1.3)
Study semester M (SD)	4.0 (1.0)	4.2 (1.3)	3.6 (1.2)	3.7 (1.7)	4.0 (1.0)
Percentage of pre-service teachers attending the academic track (%)	53.3	53.3	43.8	41.2	50.0
Pre PCK M (SD)	−1.56 † (0.63)	−1.31 † (0.49)	−1.12 † (0.64)	−1.34 † (.76)	−1.29 † (0.76)
Pre CK M (SD)	−0.78 † (0.80)	−0.70 † (0.71)	−0.61 † (0.59)	−0.73 † (0.65)	−0.60 † (0.41)
Pre PK M (SD)	0.30 † (0.42)	0.39 † (0.56)	0.38 † (0.60)	0.45 † (0.51)	0.48 † (0.41)
Pre diagnostic activities M (SD)	−2.05 † (0.47)	−2.22 † (0.94)	−1.74 † (0.68)	−1.97 † (0.65)	−2.19 † (0.78)

Note: † Values represent personal abilities from the Rasch analysis. Negative mean values for PCK, CK, and diagnostic activities indicate that the used tests were rather difficult for the pre-service teachers.

2.2. Description of the Treatments

After reviewing the literature on generic and subject-specific features of instructional quality, and including theoretical aspects that were relevant for pre-service teachers' diagnoses of subject-specific challenges in the classroom situations presented in the assessment tool *DiKoBi Assess*, content-specific aspects regarding PCK, CK, and PK were identified and summarized in texts (for an overview about the content included in the texts, see Appendix A). The texts either contained information on only one knowledge facet (treatments 1–3, see Figure 1) or represented a combined form of all three knowledge facets (treatment 4). A control group (treatment 5) did not receive any information according to the three knowledge facets. After a five minute lasting introduction into day 2 of the study, the participants of each treatment worked individually for 85 min on the associated texts. In treatments 1–4, participants were guided by identical tasks that were adapted with regard to the specific content of the texts (for an example, see Appendix B). The participants were asked to highlight important information in the texts (task 1), to complement an already outlined concept map on the basis of the highlighted information (task 2), and to apply this information by evaluating a statement or situation and providing alternatives (task 3). The tasks were constructed to consider declarative and action-related knowledge that were also part of the professional knowledge tests used [49]. However, the major share relates to declarative knowledge.

2.3. Measuring Professional Knowledge and Diagnostic Activities

2.3.1. Professional Knowledge Tests

Three paper-pencil tests for measuring PCK, CK, and PK were utilized. The tests included open-ended items (required a written response in a text field), single best answer (SBA) items (required the selection of one correct answer from a set of possible responses consisting of multiple distractors and one correct answer), and multiple true or false items (all of the possible responses had to be assessed for their validity) [78].

The PCK and CK test covered declarative and action-related knowledge about the topic “skin” (in accordance with the topic covered in the assessment tool *DiKoBi Assess* and the intervention). For example, declarative knowledge (*knowing that*) for PCK was addressed by asking for advantages and disadvantages to a specific model, which shows the structure of the human skin. Action-related knowledge in terms of *knowing when and why* was measured by asking for possible reasons students develop specific misconceptions on a specific biology topic after the learning process (cf. [49]). Both the PCK and CK test were adapted versions of the professional knowledge tests used in *ProwiN* [49,79]. The PCK test included eight open-ended items and five SBA items. Therefore, we assumed to elicit pre-service biology teachers PCK with the different items since for responding, the pre-service teachers had to draw on their individual specialized knowledge. 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 [12]. The CK test included 13 open-ended items and 15 SBA-items. Criteria for item scoring of both the PCK and CK test were provided in two separate coding manuals. Precise descriptions of the scoring process can be found in Kramer et al. [9]. To ensure objective and reliable coding, ten percent of both the PCK and CK tests 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: $ICC_{\text{absolute}}(310, 310) = 0.84, p < 0.001$; CK: $ICC_{\text{absolute}}(341, 341) = 0.97, p < 0.001$ [80].

For assessing PK, we used a short, adapted version of a paper-pencil test utilized in the *BilWiss* project covering the dimension *instruction* [40,81]. This dimension of the PK test referred to the basic dimensions of instructional quality containing declarative items about generic features such as *classroom management*, *supportive climate*, and general aspects of *cognitive activation* [11,24,28], as well as items on general pedagogical issues of teaching such as teaching methods. Since the differentiation between generic and subject-specific features of instructional quality was an important element of the video-based

assessment tool *DiKoBi Assess*, the dimension *instruction* was best suited to our construct as it referred to knowledge about generic instructional quality features, whereas the PCK-test covered knowledge about subject-specific features. For PK-measurement, participants had to answer five SBA-items and ten multiple true/false items. Item scoring followed the instructions from BilWiss [40,81]. Precise descriptions of the scoring process can be found in Kramer et al. [9].

Afterward, each knowledge test was evaluated using the Rasch partial credit model (PCM), which resulted in PCK, CK, and PK Rasch person measures for each respondent for each test instrument [82,83]. For evaluating data fit, we utilized item Outfit-MNSQ (mean-square) values, item reliability and person reliability for each test. A productive measurement is shown by item Outfit-MNSQ values below 1.5 [84]. If item reliability is high, both the range of item difficulty and the sample size can be considered as appropriate to measure the variables precisely. The person reliability is a measure of internal consistency. Person reliability is impacted by the length of the test and the range of abilities of respondents [85]. Item fit statistics of the PCK, CK, and PK test showed good fit values (see Table 2). To compare data from the identical pre- and post-tests, we anchored items from the pre-test with appropriate items from the post-test. After analyzing pre- and post-test of each knowledge facet utilizing Differential Item Functioning [82], we included 10 anchor items for the PCK test, 23 anchor items for the CK test, and 11 anchor items for the PK test. Those items, which produced a measurement bias for pre- and post-test were excluded from anchoring.

Table 2. Fit statistics of the professional knowledge tests using Rasch analysis techniques and the Rasch partial credit model.

Knowledge Facet	Number of Items	All Item Outfit-MNSQ	Person Reliability	Item Reliability
PCK pre _{anchored}	13	<1.36	0.65	0.91
PCK post	13	<1.10	0.62	0.93
CK pre _{anchored}	28	<1.37	0.70	0.95
CK post	28	<1.36	0.76	0.96
PK pre _{anchored}	15	<1.28	0.54	0.95
PK post	15	<1.18	0.60	0.95

2.3.2. Video-Based Assessment Tool *DiKoBi Assess*

To measure the three diagnostic activities (DA) *generating evidence*, *evaluating evidence*, and *drawing conclusions*, which are applicable for diagnosing instructional quality [61], we used the video-based assessment tool *DiKoBi Assess* that is embedded in an online survey platform [86]. *DiKoBi Assess* contains short staged video clips showing challenging biology classroom situations on the topic “skin”. *DiKoBi Assess* consists of six videotaped classroom situations, which represent one whole biology lesson and refer each to another subject-specific dimension of instructional quality that was found to be empirically effective for student achievement in science instruction [30,87]: (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 evaluation of these six subject-specific dimensions and the identification of subject-specific instructional quality features are applicable to any biology lesson regardless of the specific content to be taught [76]. For this study, we used two versions of the assessment tool *DiKoBi Assess*, which differed in the specific sub-theme of the embedded videos on the topic “skin”:

- *DiKoBi I Assess* (sub-theme: “skin as a sensory organ”) to assess pre-service teachers’ diagnostic activities before the intervention (diagnostic activities pre);
- *DiKoBi II Assess* (sub-theme: “protective function of the skin”) to assess pre-service teachers’ diagnostic activities after the intervention (diagnostic activities post).

However, essential is the division into the six classroom situations, which address corresponding subject-specific dimensions of instructional quality and are the same for both versions.

For each of the six classroom situations, pre-service teachers had to identify challenging aspects of the shown situation of biology instruction and reason about them by describing the identified challenging aspects (DA = *generating evidence*), by explaining (including theoretical references) why there is room for instructional improvement (DA = *evaluating evidence*), and by proposing alternative teaching strategies (DA = *drawing conclusions*) (cf. [51,60,63]). Pre-service teachers' diagnostic activities were measured in an open-ended format with short-answer items [88]. For scoring, written answers were compared with predefined sample solutions of content-related coding variables that have been compiled with regard to the literature and research results on the subject-specific dimensions of instructional quality. The content-related coding variables referred to subject-specific instructional quality features that represent challenging aspects of biology instruction (see Table 3). Results of several qualitative validation steps showed that practicing in-service biology teachers (qualified for teaching Grade 5 to 12 in German secondary schools) with an average age of 40.4 years ($SD = 9.2$) and an average teaching experience of 9.4 years ($SD = 6.9$) received the staged classroom situations as authentic and that they could identify the challenging instructional aspects sufficiently. Moreover, it was shown that the created tasks can validly measure the assumed diagnostic activities. Further information on the validation process can be found in Kramer et al. [61,76].

Table 3. Overview of the content-related coding variables of the six classroom situations, shown for the three diagnostic activities.

Classroom Situations Representing Subject-Specific Dimensions of Instructional Quality	<i>Generating Evidence</i> (13 Coding Variables)	<i>Evaluating Evidence</i> (6 Coding Variables)	<i>Drawing Conclusion</i> (12 Coding Variables)
(1) <i>Level of students' cognitive activities and creation of situational interest</i>	Level of students' cognitive activities Situational interest and motivation	Proper explanation	Level of students' cognitive activities Situational interest and motivation
(2) <i>Dealing with (specific) student ideas and errors</i>	Formative handling of student errors Identifying student error 1 Identifying student error 2	Proper explanation	Formative handling of student errors
(3) <i>Use of technical language</i>	Quantity of technical terms Quality of technical terms Explaining terms and linking with function	Proper explanation	Quantity of technical terms Quality of technical terms Explaining terms and linking with function Constructive elaboration of terms
(4) <i>Use of experiments</i>	Characteristics of scientific inquiry	Proper explanation	Characteristics of scientific inquiry
(5) <i>Use of models</i>	Elaborate model use Critical reflection	Proper explanation	Elaborate model use Critical reflection
(6) <i>Conceptual instruction</i>	Linking and back referencing Conceptual instruction	Proper explanation	Linking and back referencing Conceptual instruction

Note: The coding-variables of *generating evidence* and *drawing conclusions* are mostly identical, since both describing observations (DA = *generating evidence*) and proposing alternative strategies (DA = *drawing conclusions*) should address relevant subject-specific instructional quality features. On the other hand, there are only six coding variables for the diagnostic activity *evaluating evidence*, since—in contrast to several possible observations—the explanation for an observation can be proper or not. In other words, the given explanation must fit to the described observation.

Depending on the quality of the executed diagnostic activity (observable in not accurate or vague answers, or in more detailed, elaborated answers), Zero (0), 1, or 2 points were used for scoring answers corresponding to the diagnostic activity *generating evidence*; 0, 1, 2, or 3 points for scoring answers corresponding to *evaluating evidence*, and 0, 1, or 2 points for answers corresponding to *drawing conclusions* (cf. [9]). For a high-quality scoring of answers corresponding to the diagnostic activity *generating evidence*, it was important that the provided answer contained a systematic description of an observed challenging instructional aspect. A high-scored answer of the diagnostic activity *evaluating evidence* contained references to scientific concepts that were used to justify the claim that was made. High-quality in *drawing conclusions* became visible by specifically described alternative strategies that were derived based on the preceding steps of scientific reasoning. In the following, the scoring is exemplified for the answers of the pre-service teacher Anne to the classroom situation (4) *use of experiments*. When using experiments, biology teachers should implement the steps of scientific inquiry to foster students' scientific thinking [3,6]. However, the video shows a teaching situation (experiment on cold protection) in which the teacher disregards individual steps of scientific inquiry, such as formulating a question or generating hypotheses. Instead, the students carry out a recipe-like work instruction without having to think much. With regard to the coding variable *characteristics of scientific inquiry*, pre-service teachers' answers should address one of the aforementioned aspects. After watching the video clip, Anne gave the following description (DA = *generating evidence*): "given instruction (recipe)". Her description is very brief and does not contain many observed details. The answer is, therefore, scored 1 point. Anne justified her observation as follows (DA = *evaluating evidence*): "Students hardly learn the way of scientific inquiry". Her explanation refers to the scientific concept of scientific inquiry and is, therefore, scored 2 points. Eventually, Anne gave an alternative teaching strategy (DA = *drawing conclusions*): "Important: have students generate hypotheses, different opinions stimulate discussion, promote interest, generate excitement; experiment serves as a test of the hypothesis; discuss results; relate to practice, do fatter people then freeze less?" Anne's answer is comprehensive and refers to different aspects that are important with regard to scientific inquiry. Moreover, she proposes a transfer question which, in the sense of conceptual instruction, also represents a link to the students' everyday life. Consequently, her answer is scored 2 points. The same procedure was applied to all answers of the pre-service teachers regarding all coding variables of the three diagnostic activities.

After data collection, the scores were used to calculate person measures that represent pre-service teachers' abilities to *generate evidence*, *evaluate evidence*, or to *draw conclusions* with regard to the subject-specific classroom situations in the video-based assessment tool. Person measures were calculated by utilizing Rasch PCM. This was done separately for pre- and post-test, since the corresponding versions of the assessment tool *DiKoBi Assess* contained different sub-themes that may have had an impact on the execution of diagnostic activities. Fit statistics showed good fit values for pre- and post-test including a one-dimensional construct of diagnostic activities (diagnostic activities pre/post: 31/31 items, all item Outfit-MNSQ < 1.32/1.38, person reliability = 0.77/0.73, item reliability = 0.90/0.85; note: for diagnostic activities pre, one item in *DiKoBi I Assess* produced inestimable high values). The one-dimensional construct was used due to weak reliability values when calculating person measures separately for the three diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*.

Our video-based approach follows Kersting et al. [15], who "use video clips of authentic classroom events as prompts to elicit teachers' analyses, which are in turn assumed to draw on teachers' knowledge" (p. 571). Therefore, we assume that the videotaped classroom situations elicit pre-service biology teachers' diagnostic activities for reasoning about the biology-specific challenges and features in both versions of the assessment tool, which are in turn assumed to rely on pre-service teachers' declarative and action-related PCK (cf. [9]).

2.4. Data Analysis

First, measures of all variables (pre/post: PCK, CK, PK, diagnostic activities) were separately analyzed using the Rasch PCM [83] with the software Winsteps 3.81 [85]. Second, Pearson's correlations and descriptive results were calculated utilizing IBM SPSS Statistics (version 26) and Microsoft Excel (2010) to describe the development and intercorrelation between all variables relevant for this study. The main analysis was done in two steps. (I) To answer RQ1a and RQ1b, we ran mixed ANOVAs for PCK, CK, and PK to analyze the main and interaction effects between time and treatments. (II) To answer RQ2, we chose an analysis of covariance (ANCOVA) to analyze the effects on diagnostic activities in the post-test while controlling for diagnostic activities in the pre-test. There was homogeneity of the error variances of all variables, as assessed by Levene's test ($p > 0.05$), as well as homogeneity of covariance, as assessed by Box's test (PCK: $p = 0.53$; CK: $p = 0.57$; PK: $p = 0.53$) and homogeneity of regression slopes (diagnostic activities: $p = 0.20$).

3. Results

A descriptive overview about means and standard deviations, as well as intercorrelations between all variables, is given in Table 4. The mean values represent the average person ability from the PCM of the corresponding variables. Intercorrelations emphasized the importance of PCK since there was a low to moderate correlation between PCK and most of the pre or post-measured variables [89]. For example, PCK pre was significantly correlated with PCK post ($r = 0.57, p < 0.001$), CK post ($r = 0.30, p = 0.006$), PK pre ($r = 0.22, p = 0.048$), and diagnostic activities pre ($r = 0.33, p = 0.002$). Furthermore, the descriptive results showed that pre-service teachers' PCK, CK, and PK increased from pre to post (see Table 4). Note that the average person measures of diagnostic activities pre and post are not directly comparable because they have not been anchored, as we used two separate measurement instruments in pre and post-test (pre: DiKoBi I Assess and post: DiKoBi II Assess). However, a descriptive comparison between diagnostic activities pre and diagnostic activities post can be made based on the quality of the diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*. Table 5 shows how often each diagnostic activity was scored with 0, 1, 2, or 3 points. It is noteworthy that the total scores of *generating evidence* and *drawing conclusions* decreased from pre to post, while the total score of *evaluating evidence* increased. Furthermore, the dispersion increased from pre to post for the diagnostic activities *generating* and *evaluating evidence*. More often 0 points (low quality of diagnostic activity) but also 2 or 3 points (improved quality of diagnostic activity) were used for scoring. In contrast, fewer answers were scored with 1 point in DiKoBi II Assess. For the diagnostic activity *drawing conclusions*, the frequency of 0 points increased whereas the frequency of 1 and 2 points decreased, indicating an overall decrease in quality.

Table 4. Intercorrelations, including mean and standard deviation. Mean values of PCK, CK, and diagnostic activities are negative, because test items were rather difficult for pre-service biology teachers.

	N	M	SD	PCK Pre	PCK Post	CK Pre	CK Post	PK Pre	PK Post	Diagnostic Activities Pre	Diagnostic Activities Post
PCK pre ^a	81	−1.32	0.67	-							
PCK post ^a	81	−0.61	0.60	0.57 **	-						
CK pre ^a	81	−0.68	0.65	0.16	0.12	-					
CK post ^a	81	−0.27	0.77	0.30 **	0.38 **	0.53 **	-				
PK pre ^a	81	0.40	0.49	0.22 *	0.30 **	0.17	0.13	-			
PK post ^a	81	0.51	0.48	0.06	0.26 *	0.10	0.21	0.39 **	-		
Diagnostic activities pre ^{a,†}	81	−1.85 †	0.75 †	0.33 **	0.37 **	0.16	0.10	0.21	0.05	-	
Diagnostic activities post ^{a,†}	81	−2.00 †	0.83 †	0.15	0.27 *	0.17	0.17	0.19	0.19	0.38 **	-

^a Person abilities of variables scaled according to the PCM in Logits. † Person abilities of diagnostic activities pre and post are not anchored. Therefore, mean values pre and post are not directly comparable. ** Significant at the 0.01 level (2-tailed). * Significant at the 0.05 level (2-tailed).

Table 5. Absolute frequency of points that were used for scoring the quality of the three diagnostic activities. *N* refers to the total number of activities scored. This number is calculated from the number of participants (81) and the number of content-related coding variables for the respective diagnostic activity and is therefore the same for pre and post.

Points	Generating Evidence (N = 1053)		Evaluating Evidence (N = 486)		Drawing Conclusions (N = 972)	
	Pre	Post	Pre	Post	Pre	Post
0	796	855	345	356	658	800
1	212	139	94	54	223	143
2	45	59	47	72	91	29
3	not applicable	not applicable	0	4	not applicable	not applicable
Total score	302	257	188	210	405	201

Note: The maximum score for *generating evidence* and *drawing conclusions* was 2 points, for *evaluating evidence* it was 3 points.

Next, we ran mixed ANOVAs for PCK, CK, and PK to analyze time effect, treatment effect, and interaction effect between time and treatment (RQ1a, RQ1b). Main effects of time were found for both PCK and CK. They confirmed the positive descriptive trend of knowledge acquisition since a statistically significant increase in mean person abilities of PCK and CK was measured from pre to post. However, there was no statistically significant increase from pre- to post-measurement for PK. Furthermore, there was no statistically significant main effect of treatment or interaction between time and treatment for any knowledge facet when each treatment was considered individually (see Table 6).

Table 6. Results of mixed ANOVAs of the knowledge facets.

	df	F	<i>p</i>	Partial η^2
PCK				
- time effect	1.76	118.28	<0.001	0.61
- treatment effect	4.76	0.53	0.715	0.03
- interaction effect	4.76	1.11	0.359	0.06
CK				
- time effect	1.76	31.60	<0.001	0.29
- treatment effect	4.76	0.15	0.963	0.01
- interaction effect	4.76	2.13	0.085	0.10
PK				
- time effect	1.76	3.60	0.062	0.05
- treatment effect	4.76	0.17	0.952	0.01
- interaction effect	4.76	0.72	0.583	0.04

Note: Significant results are highlighted in bold.

However, since the maximum number of participants per treatment did not exceed 18, we merged treatments that included CK in a second step. This step was done because the *p*-value of 0.085 of the interaction effect between time and treatment for the CK group was considerably lower compared to PCK or PK. This *p*-value might indicate a possible underlying effect of the self-directed knowledge acquisition via texts in terms of CK acquisition that might not have been detectable due to the small number of participants per treatment. Since it was not possible to increase the number of participants in the overall sample for the time of the study, a fallback solution was applied: treatments containing CK and treatments not-containing CK were merged (group 1: CK-treatment and combination-treatment; group 2: PCK-treatment, PK-treatment, control group). Since our PCK-treatment did not contain CK content, we assigned the PCK-treatment to not-containing CK (group 2).

Accordingly, the PCK test was constructed in such a way that PCK could be measured as independently as possible from CK (cf. [42,79]).

However, results of mixed ANOVAs showed no statistically significant interaction effect between time and merged treatments of none of the knowledge facets either, indicating that self-directed knowledge acquisition was not effective in terms knowledge development: PCK: $F(1,79) = 0.34, p = 0.564$, partial $\eta^2 < 0.01$; CK: $F(1,79) = 0.47, p = 0.497$, partial $\eta^2 = 0.01$; PK: $F(1,79) = 1.02, p = 0.315$, partial $\eta^2 = 0.01$.

Effects of treatments on diagnostic activities were examined using an ANCOVA (RQ2). Results showed that the covariate diagnostic activities pre was significantly related to diagnostic activities post ($F(1,75) = 11.77, p = 0.001$, partial $\eta^2 = 0.14$). There was no significant effect of treatment on diagnostic activities after controlling for the effects of the covariate ($F(1,75) = 0.61, p = 0.656$, partial $\eta^2 = 0.03$), meaning that pre-service biology teachers in all treatments had equal person abilities after the intervention.

4. Discussion

This study aimed to investigate one particular way of knowledge acquisition and its effects on pre-service biology teachers' cognitive dispositions and skills. Thus, we investigated effects of self-directed knowledge acquisition via texts on the pre-service teachers' professional knowledge facets PCK, CK, and PK, as well as on their diagnostic activities as conceptualization of scientific reasoning skills in diagnostic settings. By using a video-based assessment tool to measure diagnostic activities, we contributed to situated measures of pre-service teachers' knowledge and skills.

In summary, neither PCK, CK, PK nor diagnostic activities were significantly affected by any of the study treatments. Accordingly, the research questions can be answered as follows: Self-directed knowledge acquisition via texts did not increase pre-service biology teachers' knowledge facets PCK, CK, or PK (RQ1a). Thus, no differences in the effectiveness of the intervention with respect to the three knowledge facets could be found (RQ1b). Additionally, there were no significant effects of the text-based intervention on the execution of the diagnostic activities (RQ2). In detail, we can state: The knowledge facets PCK and CK have significantly increased from pre to post; however, this increase could not be explained by the self-directed knowledge acquisition. This finding is in line with other research showing, for example, that PK could not effectively be fostered through self-directed knowledge acquisition compared to direct instruction [70]. Therefore, the methodological approach to knowledge acquisition via texts as it was implemented in the present study has not proven to be effective. However, that does not mean that self-directed knowledge acquisition via texts that represents one common learning practice at German universities is generally ineffective. Effectiveness may depend on the specific actions that are initiated by the instructional approaches. For example, Kyriakides et al. [90] found that both direct instruction and self-directed constructivist approaches can benefit student outcomes, depending on what exactly the teacher and the students do during instruction. Therefore, even though we made an effort to increase pre-service teachers' engagement with the learning material, the utilized tasks may not have been activating enough or appropriate to promote in-depth learning. Moreover, possible small effects of the self-directed knowledge acquisition via texts might have been overlaid by other effects, possibly resulting from the video-based work. However, descriptive comparisons regarding the quality of the diagnostic activities pre and post showed an increase in terms of quality that was slightly noticeable for the diagnostic activity *generating evidence* and particularly noteworthy for *evaluating evidence*. Whereas pre-service teachers' evidence evaluations in the assessment tool *DiKoBi I Assess* were often superficial and vague, their quality slightly increased in *DiKoBi II Assess* in terms of more frequent concept references and explicit linking of observations and theoretical references. This primarily indicates a potential impact of the self-directed knowledge acquisition on the diagnostic activity *evaluating evidence*. Similar findings regarding the relationship between knowledge and interpretive processes have also been described in other studies (cf. [91]). The findings on the decrease

in the quality of the diagnostic activity *drawing conclusions* from pre to post might indicate that other approaches of instructional support are necessary for the promotion of this activity, which cannot be provided via text-based instruction (that included a large amount of declarative knowledge). In addition, affective-motivational aspects have to be taken into account, because, in order to set up an alternative strategy, longer answers were required, for which the study participants might not have been motivated enough (cf. [92]). However, further research is needed to make more reliable conclusions and differentiate pathways of knowledge and skill development.

Hence, two other important questions remain whose answers contribute to the debate on measuring knowledge and skills of science teachers: First, how can the increase in knowledge in our study be explained, if not by the treatments? Second, which ways of knowledge acquisition might be more effective for university teaching? Regarding the first question, we want to refer to “the use of classroom video as a tool for bringing the central activities of teaching into the PD (professional development) setting” [93] (p. 1099). The greatest effects reported in this article are time effects of PCK-measures and, therefore, were independent of treatment. Researchers within science education already underlined that the use and prompted analysis of classroom performances challenges pre-service teachers’ thinking and thus can activate their knowledge and make it accessible [15,45,74,94]. Our hypothesis is that the increase in PCK is due to the work with the video-based assessment tool *DiKoBi Assess*. The observation and diagnosis of biology-specific classroom situations may have elicited existing subject-specific knowledge. Working on the tasks in the assessment tool *DiKoBi Assess* and engaging in scientific evidence-based reasoning on biology instruction required pre-service teachers to apply diagnostic activities. This application of diagnostic activities to a specific situation of biology instruction may have contributed to the promotion of PCK. By observing and describing challenging aspects of biology instruction (i.e., *generating evidence* that helps encoding, cf. [2]), pre-service teachers directed their focus to very specific aspects. For the evaluation of this evidence (i.e., the observed and described challenging aspects) a linkage with broader principles they represent and thus the elicitation of professional knowledge had to take place [63,73]. The scientific reasoning skill *evaluating evidence*, in particular, is considered important in order to interpret classroom interactions and inform appropriate follow-up decisions [95]. Thus, implementing opportunities in which diagnostic activities can be applied or are even fostered may in turn have an impact on pre-service teachers’ PCK. This assumption can be seen as an indication of the bidirectional relationship between knowledge and skills [96]. Consequently, the results of the study suggest that scientific reasoning about subject-specific instructional quality can potentially promote PCK and that the application of skills, such as diagnostic activities, to video-based settings thus seems to be more suitable for knowledge development than instructional support via texts. Still, a well-planned use of videos in specific teaching and learning situations is required to ensure the effectiveness of such learning opportunities (e.g., [97,98]).

Building on these thoughts, we want to emphasize the relevance of using and reflecting on practical examples for the development of pre-service teachers’ professional knowledge (cf. [94]). Although a profound declarative knowledge base is considered important (especially in terms of CK) and may still be provided via specialized texts, PCK and PK are much more action-oriented and, therefore, require other forms of knowledge acquisition (cf. [99]). Our suggestion based on the present findings is to provide learning opportunities in which pre-service teachers engage in scientific reasoning about instructional quality, for example, via video-based tools. Other researchers have already made similar suggestions. König et al. [91], for example, used video vignettes to underline the importance of practical insights into teaching to improve teachers’ general pedagogical knowledge. By setting the focus on practical scenarios, they also addressed reasoning skills as necessary components for the acquisition and transformation of knowledge. The use of video-based tools can be considered an appropriate approach to provide opportunities for the assessment and development of PCK and PK in teacher education that count as

an important part of diagnostic competences [8,29,99]. In addition, such tools could be adapted to promote other facets of knowledge. For example, videos that focus on the use of digital media in the science classroom could be utilized to promote pre-service teachers' Technological Pedagogical Content Knowledge (TPACK) that represents the knowledge necessary to effectively use technology in the classroom [100]. Moreover, a further development of digital learning environments is the use of emerging technologies such as virtual or augmented reality, which can be used to examine specific competencies of prospective teachers systematically and realistically through virtual training scenarios (cf. [101]).

Besides the use of video-based or other digital tools, other ways of supporting pre-service teachers' knowledge development have been discussed with regard to direct instruction. Effective approaches to knowledge acquisition often included an experienced lecturer in addition to text-based work in science teacher education. In contrast to static texts, a lecturer can "explicitly (address) the knowledge of students, learning and teaching in concrete content domains" [71] (p. 126), including practical examples as well, which has proven to be an effective method. Barth et al. [70] showed a positive effect of a systematic introduction to the relevant knowledge base on both the development of declarative professional knowledge and knowledge-based reasoning skills. Small effects of direct instruction on PCK were also reported by Tröbst et al. [67]. With regard to the development of PCK, the researchers found the combined instruction of the knowledge facets within professional development programs, which considered transformation processes of CK and PK during PCK-construction, to be more effective than polyvalent traditional teacher education [45,67,71]. This combined view on the acquisition of knowledge attributes a high relevance to educational training in higher education. In such trainings, practical classroom scenarios can also be videotaped and used for eliciting reasoning processes about these classroom scenarios, for which positive effects in terms of professional development could be recorded [45,93]. It is, therefore, important for pre-service teachers to take advantage of corresponding offers of universities.

5. Limitations

First of all, it should be noted that the overall sample size of the reported study was rather small, thus, resulting in an even smaller sample for the five treatments. This is accompanied by losses in terms of statistical power in our calculations. However, since the study was embedded in regular courses within a German university, the number of participants was limited and could not be easily increased up to the time of the study. In addition, extensive effects of the treatments could hardly be expected, since the intervention of the study was embedded in a seminar held once a week and lasted only 85 min. This time might not have been sufficient to produce sustainable, measurable effects. Other approaches included significantly higher intervention times with up to 250 h of professional development experiences over a two-year program [45]. At the same time, in such programs, a variety of different support activities take place to foster teachers' knowledge and skills at different levels. The advantage of the approach chosen in the present study is that by focusing on one specific activity, it is easier to investigate and monitor its effectiveness. Nevertheless, a longer intervention period might have been beneficial.

Further limitations concern the pre-service teachers' test performances. The increase in professional knowledge in terms of PCK and CK from pre to post might be the result of the pre- and post-test research design, since other experiences made during the study period or other forms of input that took place in everyday life or other seminars may have contributed to teachers' learning. Although descriptive PK-measures improved from pre to post as well, the time effect of the mixed ANOVA did not show statistical significance. One reason might be seen in the subject-specific focus of the videos, in which mainly subject-specific dimensions from biology instruction were addressed. In contrast, studies using videos that focus on general pedagogical aspects showed corresponding effects on PK [26,94]. Another reason for the lack of significant improvement in PK may also be due to the PK test used.

Test results for the pre-test were already comparatively high. Thus, the PK test did not discriminate the sample enough. Test reliabilities might be increased by utilizing a longer version of the test, or by a higher number of participants to increase variance in ability [85]. In contrast to the increase in test scores on the three knowledge facets, participants' total test scores indicating the quality of the diagnostic activities *generating evidence* and *drawing conclusions* decreased in the post-test. This could indicate a variation in the difficulty of the videos used in the two versions of the video-based assessment tool *DiKoBi Assess*. In order to investigate the dependence of the situation-specific performance of the test person on the video versions used, further analyses must be conducted that refer to the situation-specific difficulty of the video situations and thus to the situation-specificity of the participants' performances (cf. [26,102]).

Additionally, the absence of motivational variables that were not considered in this study might be seen as another potentially limiting factor. For a productive measurement, it is important that participants' desire and willingness to solve the tasks are considered and, if necessary, are scaffolded [92]. However, a supplementary analysis considering teachers' situational interest showed that controlling the variable situational interest did not change the results either. Still, an impact on diagnostic activities, particularly on *drawing conclusions*, may be possible.

A final limitation is of conceptual nature. In our study, we measured content-related knowledge facets (CK, PCK, PK). Theoretically, the different knowledge types *knowing that*, *knowing how*, *knowing when and why* can be distinguished for each of them [10]. However, the utilized PK test merely contained items covering *knowing that*. Additionally, the texts used for the intervention mainly covered declarative knowledge. Even though action-related knowledge in terms of *knowing how* and *knowing when and why* should be prompted by the tasks to be completed in the treatments, their proportion may not have been high enough to significantly influence pre-service teachers' execution of diagnostic activities. In addition, the analysis of the videos and the application of action-related knowledge therein must be considered as highly complex, which can lead to a high cognitive load, especially for learners with little prior knowledge [103]. The participants involved in this study were still at the beginning of their university education. To reduce cognitive load, short video clips were used. Furthermore, the video analysis process was already pre-structured by prompting the application of the three diagnostic activities through three individual tasks (describing, explaining, proposing alternative strategies). Moreover, the videos focused on only one dimension of subject-specific instructional quality. Nevertheless, the participants' working memory capacity may not have been sufficient to process the different information and to alternately access the corresponding knowledge types [104]. Therefore, addressing explicitly *knowing how* and *knowing when and why* is of high importance for future intervention approaches. However, they may be better accomplished through other types of instructional support than texts, e.g., through experienced lecturers.

6. Conclusions and Further Research

Finally, we want to derive implications for practice and further research. The present study provides further evidence that using video-based tools is beneficial in teacher education since these tools extend the number of practical approaches, which are provided, for example, via videos, classroom observations, or field experiences. Since the use of the video-based tool as an assessment instrument already had positive effects on pre-service teachers' professional knowledge in our study, possibly elicited by reasoning about instructional quality, it is reasonable to use the tool as a learning environment to promote pre-service biology teachers' PCK and their application of diagnostic activities even more effectively. Therefore, scientific reasoning skills should not only be investigated in terms of their relation to content knowledge of a specific discipline (cf. [4]), but also with respect to teachers' PCK. In this context, it might also be useful to use the individual activities *generating evidence*, *evaluating evidence*, and *drawing conclusions* separately for analyses instead of the one-dimensional diagnostic activities construct.

Future research could also investigate additional support in terms of scaffolding within video-based tools or simulations that might further promote the development of knowledge and diagnostic activities in order to facilitate teachers' diagnostic competences [8]. Moreover, scaffolding different types of knowledge relevant for scientific reasoning (that is procedural and epistemic knowledge for problem solving, cf. [4]) may improve scientific reasoning skills such as diagnostic activities and, thus, the development of content-related knowledge as well.

Following the demand for integrated coursework or a combined instruction of the knowledge facets, knowledge acquisition could also be addressed through direct instruction, as it is done, for example, in lectures at universities [70,71]. Therefore, further research could investigate the possible effects of an integrated presentation of the knowledge facets in lectures.

Author Contributions: Conceptualization, M.K., C.F. and B.J.N.; Formal analysis, M.K. and C.F.; Writing—Original draft preparation, M.K.; Writing—Review and editing, M.K., C.F. and B.J.N.; Project administration, B.J.N.; Funding acquisition, B.J.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG) to the project "Facilitating diagnostic competences in simulation-based learning environments in the university context" (COSIMA), NE 1196/ 8-1, FOR 2385.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Faculty of Psychology and Education of the LMU Munich in December 2017.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Information and queries on the data used can be obtained from the authors of this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Overview of the content that was covered in the texts.

Treatment	Content Included in the Texts	References
1 PCK	<ul style="list-style-type: none"> - subject-specific features of instructional quality (level of students' cognitive activities, use of originals, use of models, use of experiments and scientific working methods, use of technical language, linking and structuring of content) - lesson planning model for biology instruction considering core ideas 	[25,42,105–118]
2 CK	<ul style="list-style-type: none"> - basic functions of the skin - structure of the skin - effects of sunlight - skin as a sensory organ - perception of stimuli 	[107,119–125]
3 PK	<ul style="list-style-type: none"> - generic features of instructional quality (basic dimensions: classroom management, supportive climate, cognitive activation) - teaching methods 	[106,107,126–131]

Table A1. Cont.

Treatment	Content Included in the Texts	References
4 combination of PCK, CK, PK	<ul style="list-style-type: none"> - basic dimensions: classroom management, supportive climate, cognitive activation - lesson planning model for biology instruction considering core ideas - basic functions of the skin - structure of the skin - skin as a sensory organ 	[25,105,106,111,113–115,117,119–123,125]
5 no information (control group)	<ul style="list-style-type: none"> - reflection on the organization of the university teacher education and ideas for improvement 	-

Appendix B

Tasks:

(1) Highlight, in accordance with the key questions, important subject-specific aspects of successful teaching and learning in biology lessons in the material. Use different colours for the key questions a-c:

- a) How can cognitive activation be implemented in biology lessons?
- b) Which biology-specific features of instructional quality exist? What should be taken into account when implementing them in biology lessons?
- c) What are the main characteristics of the “Schalenmodell“ (a lesson planning model for biology instruction)?

d) In Figure 3 (see material M2)¹, assign the features mentioned in 1b by writing them down in the appropriate section of the lesson planning model.

(2) Complete the concept map below by using your highlighted information from task (1). Use your identified subject-specific aspects of successful teaching and learning in biology lessons and add at least 15 additional terms in the concept map (node terms). Make sure that the relations between the node terms are clear. To do this, link the node terms with connecting lines or arrows, and note the relationship/context in words on the corresponding connecting lines. Also use cross-links.

(3) Apply the aspects learned from the previous tasks to the following questions:

- a) Critically assess the following situation with regard to the used questions (e.g. focus question) and tasks.

Mrs. Hanna starts her lesson (on the topic blood groups) by asking the following questions: *Why do you think someone should know their blood group or why should their blood group be known? When might this be important?*

Later she presents the topic of the lesson by saying “And today we want to find out what is meant by the statement *the blood does not match*”.

During the rest of the lesson, Mrs. Hanna presents information on different types of blood cells, corresponding antigens and antibodies. The students have to repeat the information afterwards and determine the blood type of six patients.
(Example based on [115].)
- b) Describe an alternative teaching strategy in which your criticized aspects are taken into account.

Figure A1. Example of the given tasks utilized in treatment (1) PCK (translated from German). ¹ The materials used, including figures and tables, can be made available on request by the authors of this study.

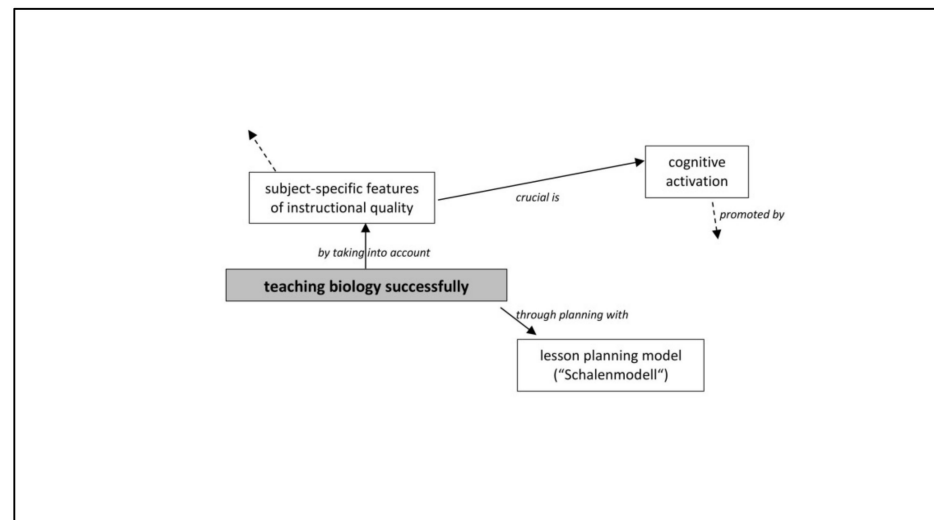


Figure A2. Outlined concept map, which is to be completed in Task 2 (translated from German).

References

1. National Research Council. *Taking Science to School: Learning and Teaching Science in Grades K–8*; National Academies Press: Washington, DC, USA, 2007.
2. Morris, B.J.; Croker, S.; Masnick, A.M.; Zimmerman, C. The Emergence of Scientific Reasoning. In *Current Topics in Children's Learning and Cognition*; Kloos, H., Morris, B.J., Amaral, J.L., Eds.; InTech: Rijeka, Croatia, 2012; pp. 61–82, ISBN 978-953-51-0855-9.
3. Krell, M.; Redman, C.; Mathesius, S.; Krüger, D.; Van Driel, J. Assessing Pre-Service Science Teachers' Scientific Reasoning Competencies. *Res. Sci. Educ.* **2020**, *50*, 2305–2329. [[CrossRef](#)]
4. Kind, P.; Osborne, J. Styles of Scientific Reasoning: A Cultural Rationale for Science Education? *Sci. Educ.* **2017**, *101*, 8–31. [[CrossRef](#)]
5. Nowak, K.H.; Nehring, A.; Tiemann, R.; Zu Belzen, A.U. Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *J. Biol. Educ.* **2013**, *47*, 182–188. [[CrossRef](#)]
6. Mayer, J. Erkenntnisgewinnung als wissenschaftliches Problemlösen (Inquiry as scientific problem solving). In *Theorien in der Biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden (Theories in Biology Education Research. A Handbook for Pre-Service Teachers and Doctoral Students)*; Krüger, D., Vogt, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 177–186.
7. Helmke, A.; Lenske, G. Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung (Diagnosis of Classroom Instruction from Different Perspectives as a Prerequisite for Improving Teaching and Learning). *Beitr. Lehrerbildung* **2013**, *31*, 214–233.
8. Heitzmann, N.; Seidel, T.; Hetmanek, A.; Wecker, C.; Fischer, M.R.; Ufer, S.; Schmidmaier, R.; Neuhaus, B.; Siebeck, M.; Stürmer, K.; et al. Facilitating Diagnostic Competences in Simulations in Higher Education A Framework and a Research Agenda. *Front. Learn. Res.* **2019**, *7*, 1–24. [[CrossRef](#)]
9. Kramer, M.; Förtsch, C.; Boone, W.J.; Seidel, T.; Neuhaus, B.J. Investigating Pre-Service Biology Teachers' Diagnostic Competences: Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy. *Educ. Sci.* **2021**, *11*, 89. [[CrossRef](#)]
10. Förtsch, C.; Sommerhoff, D.; Fischer, F.; Fischer, M.R.; Girwidz, R.; Obersteiner, A.; Reiss, K.; Stürmer, K.; Siebeck, M.; Schmidmaier, R.; et al. Systematizing Professional Knowledge of Medical Doctors and Teachers: Development of an Interdisciplinary Framework in the Context of Diagnostic Competences. *Educ. Sci.* **2018**, *8*, 207. [[CrossRef](#)]
11. Baumert, J.; Kunter, M.; Blum, W.; Brunner, M.; Voss, T.; Jordan, A.; Klusmann, U.; Krauss, S.; Neubrand, M.; Tsai, Y.-M. Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *Am. Educ. Res. J.* **2010**, *47*, 133–180. [[CrossRef](#)]
12. Tepner, O.; Borowski, A.; Dollny, S.; Fischer, H.E.; Jüttner, M.; Kirschner, S.; Leutner, D.; Neuhaus, B.J.; Sandmann, A.; Sumfleth, E.; et al. Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften (Item Development Model for Assessing Professional Knowledge of Science Teachers). *ZfDN* **2012**, *18*.
13. TEDS-M 2008. *Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im Internationalen Vergleich (TEDS-M 2008. Professional Competence and Learning Opportunities of Prospective Secondary Mathematics Teachers in International Comparison)*; Blömeke, S., Kaiser, G., Lehmann, R., Eds.; Waxmann: Münster, Germany, 2010.
14. Ball, D.L.; Thames, M.H.; Phelps, G. Content Knowledge for Teaching. *J. Teach. Educ.* **2008**, *59*, 389–407. [[CrossRef](#)]
15. Kersting, N.B.; Givvin, K.B.; Thompson, B.J.; Santagata, R.; Stigler, J.W. Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *Am. Educ. Res. J.* **2012**, *49*, 568–589. [[CrossRef](#)]
16. Hoth, J.; Döhrmann, M.; Kaiser, G.; Busse, A.; König, J.; Blömeke, S. Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM Math. Educ.* **2016**, *48*, 41–53. [[CrossRef](#)]

17. Blömeke, S.; Busse, A.; Kaiser, G.; König, J.; Suhl, U. The relation between content-specific and general teacher knowledge and skills. *Teach. Teach. Educ.* **2016**, *56*, 35–46. [[CrossRef](#)]
18. Blömeke, S.; Gustafsson, J.-E.; Shavelson, R.J. Beyond Dichotomies. *Z. Psychol.* **2015**, *223*, 3–13. [[CrossRef](#)]
19. Boyatzis, R.E. *The Competent Manager*; Wiley: New York, NY, USA, 1982.
20. Kaiser, G.; Blömeke, S.; König, J.; Busse, A.; Döhrmann, M.; Hoth, J. Erratum to: Professional competencies of (prospective) mathematics teachers—Cognitive versus situated approaches. *Educ. Stud. Math.* **2016**, *94*, 183–184. [[CrossRef](#)]
21. Depaepe, F.; Verschaffel, L.; Kelchtermans, G. Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teach. Teach. Educ.* **2013**, *34*, 12–25. [[CrossRef](#)]
22. Zimmerman, C. The Development of Scientific Reasoning Skills. *Dev. Rev.* **2000**, *20*, 99–149. [[CrossRef](#)]
23. Seidel, T.; Shavelson, R.J. Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Rev. Educ. Res.* **2007**, *77*, 454–499. [[CrossRef](#)]
24. Lipowsky, F.; Rakoczy, K.; Pauli, C.; Drollinger-Vetter, B.; Klieme, E.; Reusser, K. Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learn. Instr.* **2009**, *19*, 527–537. [[CrossRef](#)]
25. Wüsten, S. Allgemeine und Fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie: Eine Video- und Interventionstudie (General and Subject-Specific Instructional Quality in the Subject Biology: A Video and Intervention Study). Ph.D. Thesis, Universität Duisburg-Essen, Duisburg, Germany, 2010.
26. König, J.; Kramer, C. Teacher professional knowledge and classroom management: On the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM Math. Educ.* **2015**, *48*, 139–151. [[CrossRef](#)]
27. Praetorius, A.-K.; Klieme, E.; Herbert, B.; Pinger, P. Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Math. Educ.* **2018**, *50*, 407–426. [[CrossRef](#)]
28. Klieme, E.; Schümer, G.; Knoll, S. Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung (Mathematics Instruction in Secondary Education: Task Culture and Instructional Processes). In *TIMMS—Impulse für Schule und Unterricht (TIMSS—Impetus for School and Teaching)*; Bundesministerium für Bildung und Forschung, Ed.; Bundesministerium für Bildung und Forschung (BMBF): Bonn, Germany, 2001; pp. 43–57.
29. Meschede, N.; Fiebranz, A.; Möller, K.; Steffensky, M. Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teach. Teach. Educ.* **2017**, *66*, 158–170. [[CrossRef](#)]
30. Schlesinger, L.; Jentsch, A. Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM Math. Educ.* **2016**, *48*, 29–40. [[CrossRef](#)]
31. Lai, M.K.; Schildkamp, K. In-service Teacher Professional Learning: Use of Assessment in Data-based Decision-making. In *Handbook of Human and Social Conditions in Assessment*; Brown, G., Ed.; Routledge: London, UK, 2016; pp. 77–94, ISBN 9781315749136.
32. Weinert, F.E.; Schrader, F.-W.; Helmke, A. Educational expertise. Closing the gap between educational research and classroom practice. *Sch. Psychol. Int.* **1990**, *11*, 163–180. [[CrossRef](#)]
33. Baumert, J.; Kunter, M. The COACTIV Model of Teachers' Professional Competence. In *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*; Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M., Eds.; Springer: New York, NY, USA, 2013; pp. 25–48, ISBN 978-1-4614-5148-8.
34. Fischer, F.; Kollar, I.; Ufer, S.; Sodian, B.; Hussmann, H.; Pekrun, R.; Neuhaus, B.; Dorner, B.; Pankofer, S.; Fischer, M.; et al. Frontline Learning Research. *Front. Learn. Res.* **2014**, *2*, 28–45. [[CrossRef](#)]
35. Paris, S.G.; Lipson, M.Y.; Wixson, K.K. Becoming a strategic reader. *Contemp. Educ. Psychol.* **1983**, *8*, 293–316. [[CrossRef](#)]
36. Shulman, L. Knowledge and Teaching: Foundations of the New Reform. *Harv. Educ. Rev.* **1987**, *57*, 1–23. [[CrossRef](#)]
37. Shulman, L.S. Those who understand: Knowledge growth in teaching. *Educ. Res.* **1986**, *15*, 4–14. [[CrossRef](#)]
38. Voss, T.; Kunter, M.; Baumert, J. Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *J. Educ. Psychol.* **2011**, *103*, 952–969. [[CrossRef](#)]
39. Fischer, H.E.; Borowski, A.; Tepner, O. Professional Knowledge of Science Teachers. In *Second International Handbook of Science Education*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2011; pp. 435–448.
40. Kunina-Habenicht, O.; Maurer, C.; Wolf, K.; Holzberger, D.; Schmidt, M.; Dicke, T.; Teuber, Z.; Koc-Januchta, M.; Lohse-Bossenz, H.; Leutner, D.; et al. 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* **2020**, *66*, 80–92. [[CrossRef](#)]
41. Gess-Newsome, J. 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*; Berry, A., Friedrichsen, P., Loughran, J., Eds.; Routledge: New York, NY, USA, 2015; pp. 28–42.
42. Förtsch, C.; Werner, S.; von Kotzebue, L.; Neuhaus, B.J. Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *Int. J. Sci. Educ.* **2016**, *38*, 2642–2666. [[CrossRef](#)]
43. Schmelzing, S.; Van Driel, J.; Jüttner, M.; Brandenbusch, S.; Sandmann, A.; Neuhaus, B.J. 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.* **2013**, *11*, 1369–1390. [[CrossRef](#)]
44. Förtsch, S.; Förtsch, C.; von Kotzebue, L.; Neuhaus, B.J. Effects of Teachers' Professional Knowledge and Their Use of Three-Dimensional Physical Models in Biology Lessons on Students' Achievement. *Educ. Sci.* **2018**, *8*, 118. [[CrossRef](#)]

45. Gess-Newsome, J.; Taylor, J.A.; Carlson, J.; Gardner, A.L.; Wilson, C.D.; Stuhlsatz, M.A.M. Teacher pedagogical content knowledge, practice, and student achievement. *Int. J. Sci. Educ.* **2017**, *41*, 944–963. [[CrossRef](#)]
46. Alonzo, A.C.; Berry, A.; Nilsson, P. Unpacking the Complexity of Science Teachers' PCK in Action: Enacted and Personal PCK. In *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*; Hume, A., Cooper, R., Borowski, A., Eds.; Springer: Singapore, 2019; pp. 271–286.
47. Magnusson, S.; Krajcik, J.; Borko, H. Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching. In *Science & Technology Education Library*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2006; pp. 95–132.
48. Carlson, J.; Daehler, K.R.; Alonzo, A.C.; Barendsen, E.; Berry, A.; Borowski, A.; Carpendale, J.; Chan, K.K.H.; Cooper, R.; Friedrichsen, P.; et al. The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*; Springer International Publishing: Singapore, 2019; pp. 77–94.
49. Jüttner, M.; Boone, W.J.; Park, S.; Neuhaus, B.J. Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educ. Asse. Eval. Acc.* **2013**, *25*, 45–67. [[CrossRef](#)]
50. Shavelson, R.J. On the measurement of competency. *Empir. Res. Vocat. Educ. Train.* **2010**, *2*, 41–63. [[CrossRef](#)]
51. Kaiser, G.; Busse, A.; Hoth, J.; König, J.; Blömeke, S. About the Complexities of Video-Based Assessments: Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *Int. J. Sci. Math. Educ.* **2015**, *13*, 369–387. [[CrossRef](#)]
52. Brown, N.J.S.; Furtak, E.M.; Timms, M.; Nagashima, S.O.; Wilson, M. The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning. *Educ. Assess.* **2010**, *15*, 123–141. [[CrossRef](#)]
53. Dorfner, T.; Förtsch, C.; Germ, M.; Neuhaus, B.J. Biology instruction using a generic framework of scientific reasoning and argumentation. *Teach. Teach. Educ.* **2018**, *75*, 232–243. [[CrossRef](#)]
54. European Commission. *Science Education for Responsible Citizenship: Report to the European Commission of the Expert Group on Science Education*; Publications Office of the European Union: Luxembourg, 2015; ISBN 9789279436369.
55. Wildgans-Lang, A.; Scheuerer, S.; Obersteiner, A.; Fischer, F.; Reiss, K. Analyzing prospective mathematics teachers' diagnostic processes in a simulated environment. *ZDM Math. Educ.* **2020**, *52*, 241–254. [[CrossRef](#)]
56. Loibl, K.; Leuders, T.; Dörfler, T. A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teach. Teach. Educ.* **2020**, *91*, 103059. [[CrossRef](#)]
57. Ostermann, A.; Leuders, T.; Nückles, M. Improving the judgment of task difficulties: Prospective teachers' diagnostic competence in the area of functions and graphs. *J. Math. Teach. Educ.* **2017**, *21*, 579–605. [[CrossRef](#)]
58. Ohle, A.; McElvany, N.; Horz, H.; Ullrich, M. Text-picture integration—Teachers' attitudes, motivation and self-related cognitions in diagnostics. *J. Educ. Res. Online* **2015**, *7*, 11–33.
59. Kim, S.; Rehder, B. How prior knowledge affects selective attention during category learning: An eyetracking study. *Mem. Cogn.* **2010**, *39*, 649–665. [[CrossRef](#)]
60. Bauer, E.; Fischer, F.; Kiesewetter, J.; Shaffer, D.W.; Fischer, M.R.; Zottmann, J.M.; Sailer, M. Diagnostic Activities and Diagnostic Practices in Medical Education and Teacher Education: An Interdisciplinary Comparison. *Front. Psychol.* **2020**, *11*, 562665. [[CrossRef](#)]
61. Kramer, M.; Förtsch, C.; Seidel, T.; Neuhaus, B.J. Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Stud. Educ. Eval.* **2021**, *28*. [[CrossRef](#)]
62. Sherin, M.G.; Van Es, E.A. Effects of Video Club Participation on Teachers' Professional Vision. *J. Teach. Educ.* **2008**, *60*, 20–37. [[CrossRef](#)]
63. Seidel, T.; Stürmer, K. Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *Am. Educ. Res. J.* **2014**, *51*, 739–771. [[CrossRef](#)]
64. Hetmanek, A.; Engelmann, K.; Opitz, A.; Fischer, F. *Beyond Intelligence and Domain Knowledge: Scientific Reasoning and Argumentation as a Set of Cross-Domain Skills. Scientific Reasoning and Argumentation: The roles of Domain-Specific and Domain-General Knowledge*; Routledge: New York, NY, USA, 2018.
65. Krell, M.; Reinisch, B.; Krüger, D. Analyzing Students' Understanding of Models and Modeling Referring to the Disciplines Biology, Chemistry, and Physics. *Res. Sci. Educ.* **2015**, *45*, 367–393. [[CrossRef](#)]
66. Klemm, J.; Flores, P.; Sodian, B.; Neuhaus, B.J. Scientific Reasoning in Biology – the Impact of Domain-General and Domain-Specific Concepts on Children's Observation Competency. *Front. Psychol.* **2020**, *11*, 53. [[CrossRef](#)]
67. Tröbst, S.; Kleickmann, T.; Depaepe, F.; Heinze, A.; Kunter, M. Effects of instruction on pedagogical content knowledge about fractions in sixth-grade mathematics on content knowledge and pedagogical knowledge. *Unterrichtswissenschaft* **2019**, *47*, 79–97. [[CrossRef](#)]
68. Meier, M.; Grospietsch, F.; Mayer, J. Vernetzung von Wissensfacetten professioneller Handlungskompetenz in hochschuldidaktischen Lehr-Lernsettings (Linking of knowledge facets of professional acting in teaching-learning settings of higher education). In *Kohärenz in der Universitären Lehrerbildung*; Glowinski, I., Borowski, A., Gillen, J., Schanze, S., Von Meien, J., Eds.; Universitätsverlag Potsdam: Potsdam, Germany, 2018; pp. 143–178.

69. Gruber, H.; Mandl, H.; Renkl, A. Was lernen wir in der Schule und Hochschule: Träges Wissen? (What do we learn in school and college: Inert knowledge?). In *Die Kluft zwischen Wissen und Handeln. Empirische und Theoretische Lösungsansätze (The Gap Between Knowledge and Action. Empirical and Theoretical Approaches)*; Mandl, H., Gerstenmaier, J., Eds.; Hogrefe: Göttingen, Germany, 2000; pp. 139–156.
70. Barth, V.L.; Piwowar, V.; Kumschick, I.R.; Ophardt, D.; Thiel, F. The impact of direct instruction in a problem-based learning setting. Effects of a video-based training program to foster preservice teachers' professional vision of critical incidents in the classroom. *Int. J. Educ. Res.* **2019**, *95*, 1–12. [CrossRef]
71. Kleickmann, T.; Tröbst, S.; Heinze, A.; Bernholt, A.; Rink, R.; Kunter, M. Teacher Knowledge Experiment: Conditions of the Development of Pedagogical Content Knowledge. In *Competence Assessment in Education: Research, Models and Instruments*; Leutner, D., Fleischer, J., Grünkorn, J., Klieme, E., Eds.; Springer: Cham, Switzerland, 2017; pp. 111–129, ISBN 978-3-319-50028-7.
72. Smit, R.; Weitzel, H.; Blank, R.; Rietz, F.; Tardent, J.; Robin, N. Interplay of secondary pre-service teacher content knowledge (CK), pedagogical content knowledge (PCK) and attitudes regarding scientific inquiry teaching within teacher training. *Res. Sci. Technol. Educ.* **2017**, *35*, 477–499. [CrossRef]
73. Van Es, E.A.; Sherin, M.G. Mathematics teachers' "learning to notice" in the context of a video club. *Teach. Teach. Educ.* **2008**, *24*, 244–276. [CrossRef]
74. Santagata, R.; Guarino, J. Using video to teach future teachers to learn from teaching. *ZDM Math. Educ.* **2011**, *43*, 133–145. [CrossRef]
75. Furtak, E.M.; Bakeman, R.; Buell, J. Developing knowledge-in-action with a learning progression: Sequential analysis of teachers' questions and responses to student ideas. *Teach. Teach. Educ.* **2018**, *76*, 267–282. [CrossRef]
76. Kramer, M.; Förtsch, C.; Stürmer, J.; Förtsch, S.; Seidel, T.; Neuhaus, B.J. Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education* **2020**, *7*. [CrossRef]
77. Cortina, K.S.; Thames, M.H. Teacher Education in Germany. In *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project*; Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M., Eds.; Springer: New York, NY, USA, 2013; pp. 49–62, ISBN 978-1-4614-5148-8.
78. Campbell, D. How to write good multiple-choice questions. *J. Paediatr. Child Health* **2011**, *47*, 322–325. [CrossRef]
79. Jüttner, M.; Neuhaus, B.J. Development of Items for a Pedagogical Content Knowledge Test Based on Empirical Analysis of Pupils' Errors. *Int. J. Sci. Educ.* **2012**, *34*, 1125–1143. [CrossRef]
80. Wirtz, M.A.; Caspar, F. *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen (Interrateragreement and Interraterreliability. Methods for Determining and Improving the Reliability of Assessments via Category Systems and Rating Scales)*; Hogrefe: Göttingen, Germany, 2002; ISBN 3801716465.
81. Kunter, M.; Leutner, D.; Terhart, E.; Baumert, J. Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung (BilWiss) (Broad Pedagogical Knowledge and the Development of Professional Competence in Teacher Education (BilWiss)) (Version 5) (Data Set). Available online: <https://www.iqb.hu-berlin.de/fdz/studies/BilWiss> (accessed on 10 February 2021).
82. Boone, W.J.; Staver, J.R.; Yale, M.S. *Rasch Analysis in the Human Sciences*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2014.
83. Bond, T.G.; Fox, C.M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. *J. Educ. Meas.* **2003**, *40*, 185–187. [CrossRef]
84. Wright, B.D.; Linacre, J.M. Reasonable mean-square fit values. *Rasch Meas. Trans.* **1996**, *2*, 370.
85. Linacre, J.M. A User's Guide to Winsteps® Ministeps Rasch-Model Computer Programs: Program Manual 3.81.0. Available online: <http://www.winsteps.com/manuals.htm> (accessed on 10 February 2021).
86. Questback GmbH. *EFS Survey*; Questback GmbH: Köln, Germany, 2018.
87. Dorfner, T.; Förtsch, C.; Neuhaus, B.J. Die methodische und inhaltliche Ausrichtung quantitativer Videostudien zur Unterricht-squalität im mathematisch-naturwissenschaftlichen Unterricht: Ein Review [The methodical and content-related orientation of quantitative video studies on instructional quality in mathematics and science education. A review]. *ZfDN* **2017**, *23*, 261–285. [CrossRef]
88. Kersting, N.B.; Sherin, B.L.; Stigler, J.W. Automated Scoring of Teachers' Open-Ended Responses to Video Prompts. *Educ. Psychol. Meas.* **2014**, *74*, 950–974. [CrossRef]
89. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum: Hillsdale, NJ, USA, 1988.
90. Kyriakides, L.; Christoforou, C.; Charalambous, C.Y. What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teach. Teach. Educ.* **2013**, *36*, 143–152. [CrossRef]
91. König, J.; Blömeke, S.; Klein, P.; Suhl, U.; Busse, A.; Kaiser, G. Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teach. Teach. Educ.* **2014**, *38*, 76–88. [CrossRef]
92. Belland, B.R.; Kim, C.; Hannafin, M.J. A Framework for Designing Scaffolds That Improve Motivation and Cognition. *Educ. Psychol.* **2013**, *48*, 243–270. [CrossRef] [PubMed]
93. Osborne, J.F.; Borko, H.; Fishman, E.; Zaccarelli, F.G.; Berson, E.; Busch, K.C.; Reigh, E.; Tseng, A. Impacts of a Practice-Based Professional Development Program on Elementary Teachers' Facilitation of and Student Engagement with Scientific Argumentation. *Am. Educ. Res. J.* **2019**, *56*, 1067–1112. [CrossRef]

94. Stuermer, K.; Könings, K.D.; Seidel, T. Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *Br. J. Educ. Psychol.* **2012**, *83*, 467–483. [CrossRef]
95. Van Es, E.A.; Sherin, M.G. Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *J. Technol. Teach.* **2002**, *10*, 571–596.
96. Santagata, R.; Yeh, C. The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM Math. Educ* **2016**, *48*, 153–165. [CrossRef]
97. Gaudin, C.; Chaliès, S. Video viewing in teacher education and professional development: A literature review. *Educ. Res. Rev.* **2015**, *16*, 41–67. [CrossRef]
98. Blomberg, G.; Renkl, A.; Gamoran Sherin, M.; Borko, H.; Seidel, T. Five research-based heuristics for using video in pre-service teacher education. *J. Educ. Res. Online* **2013**, *5*, 90–114.
99. Kaiser, G.; Blömeke, S.; Busse, A.; Döhrmann, M.; König, J. Professional knowledge of (prospective) Mathematics teachers—Its structure and development. In *Proceedings of the Joint Meeting of PME 38 and PME-NA 36*; Liljedahl, P., Nicol, C., Oesterle, S., Allan, D., Eds.; PME: Vancouver, BC, Canada, 2014; pp. 35–50.
100. Mishra, P.; Koehler, M.J. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teach. Coll. Rec.* **2006**, *108*, 1017–1054. [CrossRef]
101. Huang, Y.; Richter, E.; Kleickmann, T.; Wiepke, A.; Richter, D. Classroom complexity affects student teachers' behavior in a VR classroom. *Comput. Educ.* **2021**, *163*. [CrossRef]
102. Blömeke, S.; König, J.; Suhl, U.; Hoth, J.; Döhrmann, M. Wie situationsbezogen ist die Kompetenz von Lehrkräften? Zur Generalisierbarkeit der Ergebnisse von videobasierten Performanztests (To What Extent Is Teacher Competence Situation-Related? On the generalizability of the results of video-based performance tests). *Z. Pädagog.* **2015**, *61*, 310–327.
103. Harr, N.; Eichler, A.; Renkl, A. Integrated learning: Ways of fostering the applicability of teachers' pedagogical and psychological knowledge. *Front. Psychol.* **2015**, *6*, 1–16. [CrossRef]
104. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer: Berlin/Heidelberg, Germany, 2011.
105. Werner, S. Zusammenhänge zwischen dem Fachspezifischen Professionswissen einer Lehrkraft, dessen Unterrichtsgestaltung und Schülervariablen am Beispiel eines Elaborierten Modelleinsatzes (Relationships between a Teacher's Subject-Specific Professional Knowledge, their Instructional Design, and Student Variables Using the Example of an Elaborate Model Use). Ph.D. Thesis, Ludwig-Maximilians Universität, München, Germany, 2016.
106. Seidel, T.; Schindler, A.-K. Klassenführung. In *Handwörterbuch Pädagogische Psychologie (Handbook of Educational Psychology)*, 5th ed.; Rost, D.H., Sparfeldt, J.R., Buch, S.R., Eds.; Beltz: Weinheim, Germany, 2018; pp. 328–336, ISBN 978-3-621-28297-0.
107. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Standards für die Lehrerbildung—Bildungswissenschaften. Available online: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf (accessed on 16 October 2020).
108. Klieme, E.; Pauli, C.; Reusser, K. The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Mathematics Classrooms. In *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*; Janik, T., Seidel, T., Eds.; Waxmann: Münster, Germany, 2009; pp. 137–160.
109. Kattmann, U. Modelle (Models). In *Fachdidaktik Biologie*, 8th ed.; Gropengießer, H., Kattmann, U., Eds.; Aulis: Köln, Germany, 2008; pp. 330–339.
110. Jatzwauk, P.; Rumann, S.; Sandmann, A. Der Einfluss des Aufgabeneinsatzes im Biologieunterricht auf die Lernleistung der Schüler—Ergebnisse einer Videostudie (The effect of usage of tasks in biology education on learning performance—A video study). *ZfDN* **2008**, *14*, 263–282.
111. Haugwitz, M. Kontextorientiertes Lernen und Concept Mapping im Fach Biologie. Eine experimentelle Untersuchung zum Einfluss auf Interesse und Leistung unter Berücksichtigung von Moderationseffekten individueller Voraussetzungen beim kooperativen Lernen (Context-Oriented Learning and Concept Mapping in Biology. An Experimental Investigation of the Influence on Interest and Achievement Considering Moderation Effects of Individual Prerequisites in Cooperative Learning). Ph.D. Thesis, Universität Duisburg-Essen, Duisburg, Germany, 2009.
112. Förtsch, C.; Dorfner, T.; Baumgartner, J.; Werner, S.; von Kotzebue, L.; Neuhaus, B.J. Fostering Students' Conceptual Knowledge in Biology in the Context of German National Education Standards. *Res. Sci. Educ.* **2020**, *50*, 739–771. [CrossRef]
113. Dorfner, T.; Förtsch, C.; Neuhaus, B.J. Use of Technical Terms in German Biology Lessons and its Effects on Students' Conceptual Learning. *Res. Sci. Technol. Educ.* **2019**, *38*, 227–251. [CrossRef]
114. Dorfner, T.; Förtsch, C.; Neuhaus, B.J. Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction. *Learn. Instr.* **2018**, *56*, 42–53. [CrossRef]
115. Nawani, J.; Kotzebue, L.; Rixius, J.; Graml, M.; Neuhaus, B.J. Teachers' Use of Focus Questions in German Biology Classrooms: A Video-based Naturalistic Study. *Int. J. Sci. Math. Educ.* **2017**, *95*, 639. [CrossRef]
116. Förtsch, C.; Werner, S.; Dorfner, T.; von Kotzebue, L.; Neuhaus, B.J. Effects of Cognitive Activation in Biology Lessons on Students' Situational Interest and Achievement. *Res. Sci. Educ.* **2017**, *47*, 559–578. [CrossRef]
117. Wadouh, J.; Liu, N.; Sandmann, A.; Neuhaus, B.J. The Effect of Knowledge linking Levels in Biology Lessons upon Students' Knowledge Structure. *Int. J. Sci. Math. Educ.* **2014**, *12*, 25–47. [CrossRef]
118. Neuhaus, B.J.; Nachreiner, K.; Oberbei, I.; Spangler, M. Basiskonzepte zur Planung von Biologieunterricht. Ein Gedankenspiel [Core ideas for biology lesson planning. A mental game]. *MNU (Math.-Nat. Unterr.)* **2014**, *67*, 160–163.

119. Thews, G.; Mutschler, E.; Vaupel, P. *Anatomie Physiologie Pathophysiologie des Menschen. Ein Lehrbuch für Pharmazeuten und Biologen (Anatomy Physiology Pathophysiology of the Human Being. A Textbook for Pharmacists and Biologists)*; Wissenschaftliche Verlagsgesellschaft mbH.: Stuttgart, Germany, 1980.
120. Purves, W.K.; Sadava, D.; Orians, G.H.; Heller, H.C. *Biologie (Biology)*, 7th ed.; Elsevier: Munich, Germany, 2004.
121. Müller, W.; Frings, S. *Tier- und Humanphysiologie. Eine Einführung (Animal and Human Physiology. An Introduction)*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2007; ISBN 9783540327332.
122. Moyes, C.D.; Schulte, P.M. *Tierphysiologie (Animal Physiology)*, 1st ed.; Pearson Studium: München, Germany, 2008; ISBN 9783827372703.
123. Mörike, K.D.; Betz, E.; Mergenthaler, W. *Biologie des Menschen (Human Biology)*, 15th ed.; Nikol: Hamburg, Germany, 2007; ISBN 9783937872551.
124. Hildebrand, M.; Goslow, G.E. *Vergleichende und Funktionelle Anatomie der Wirbeltiere (Comparative and Functional Anatomy of Vertebrates)*; Springer: Berlin/Heidelberg, Germany, 2004; ISBN 9783642189517.
125. Hickman, C.P.; Roberts, L.S.; Larson, A.; L'Anson, H.; Eisenhour, D.J. *Zoologie (Zoology)*, 13th ed.; Pearson Education: München, Germany, 2008; ISBN 3827372658.
126. Koran, M.L.; Koran, J.J. Aptitude-treatment interaction research in science education. *J. Res. Sci. Teach.* **1984**, *21*, 793–808. [[CrossRef](#)]
127. *Biologieunterricht Heute. Eine Moderne Fachdidaktik (Biology Classes Today. A Modern View on Biology Education)*, 12th ed.; Killermann, W., Hiering, P., Starosta, B., Eds.; Auer: Donauwörth, Germany, 2008.
128. Helmke, A.; Brühwiler, C. Unterrichtsqualität (Instructional Quality). In *Handwörterbuch Pädagogische Psychologie (Handbook of Educational Psychology)*, 5th ed.; Rost, D.H., Sparfeldt, J.R., Buch, S.R., Eds.; Beltz: Weinheim, Germany, 2018; pp. 860–869, ISBN 978-3-621-28297-0.
129. Helmke, A. *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts (Quality of Teaching and Teacher Professionalism. Diagnosis, Evaluation and Improvement of teaching)*, 5th ed.; Klett-Kallmeyer: Seelze, Germany, 2014.
130. Helmke, A. Aktive Lernzeit optimieren—Was wissen wir über effiziente Klassenführung? (Optimizing Time on Task—What do we know about effective classroom management?). *Pädagogik* **2007**, *59*, 44–49.
131. Eder, F. Schul- und Klassenklima. In *Handwörterbuch Pädagogische Psychologie (Handbook of Educational Psychology)*, 5th ed.; Rost, D.H., Sparfeldt, J.R., Buch, S.R., Eds.; Beltz: Weinheim, Germany, 2018; pp. 696–707, ISBN 978-3-621-28297-0.