Article

Investigating Pre-Service Biology Teachers' Diagnostic Competences: Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy

Maria Kramer [1],*[ID], Christian Förtsch [1][ID], William J. Boone [2][ID], Tina Seidel [3] and Birgit J. Neuhaus [1]

[1] 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.)
[2] Department of Educational Psychology, Miami University, Oxford, OH 45056, USA; boonewj@miamioh.edu
[3] School of Education, Technical University of Munich, 80335 Munich, Germany; tina.seidel@tum.de
* Correspondence: maria.kramer@biologie.uni-muenchen.de

Abstract: Teachers' diagnostic competences are essential with respect to student achievement, classroom assessment, and instructional quality. Important components of diagnostic competences are teachers' professional knowledge including content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK), their diagnostic activities as a specification of situation-specific skills, and diagnostic accuracy. Accuracy is determined by comparing a teacher's observation of classroom incidents with subject-specific challenges to be identified from scripted instructional situations. To approximate diagnostic situations close to real-life, the assessment of science teachers' diagnostic competences requires a situated context that was provided through videotaped classroom situations in this study. We investigated the relationship between professional knowledge (PCK, CK, PK) of 186 pre-service biology teachers, their diagnostic activities, and diagnostic accuracy measured with the video-based assessment tool DiKoBi *Assess*. Results of path analyses utilizing Rasch measures showed that both PCK and PK were statistically significantly related to pre-service teachers' diagnostic activities. Additionally, biology teachers' PCK was positively related to diagnostic accuracy. Considering higher effect sizes of PCK compared to PK, the findings support previous findings indicating the importance of PCK, thus demonstrating its importance in the context of subject-specific diagnosis as well.

Keywords: diagnostic competences; teacher professional knowledge; diagnostic activities; diagnostic accuracy; pedagogical content knowledge; video-based assessment; biology education

1. Introduction

In order to make efficient decisions during classroom instruction, teachers' abilities to identify and interpret relevant situations and events that influence student learning have been described as an important part of teachers' professional competence (e.g., [1,2]). Similar to the processes that a medical person performs (identifying and interpreting symptoms of a patient to decide on how to treat the patient best), the ability to assess classroom situations and events (e.g., how to implement an experiment in science instruction best, or how to deal with student misconceptions) in order to adapt instruction can be considered in the context of diagnostic competences [3,4]. Therefore, the consideration of diagnostic competences is a critical element in teacher education, which is significantly important with respect to student achievement, classroom assessment, and instructional quality [5–7]. Its importance is also stated in German standards of teacher education [8]. At this point, it is important to remark that in German research, diagnostic competences are used largely synonymous with assessment competences [4,9]. For reasons of consistency, we primarily use the term diagnostic competences instead of assessment competences throughout this article. The plural is used to indicate that there is no global construct of diagnostic competence but that its conceptualization depends on the study's specific focus [10,11].

Nonetheless, courses and practicums that promote diagnostic competences vary between universities [12,13]. Likewise, approaches to support diagnostic competences vary based on different diagnostic contexts (e.g., assessing learning outcomes, diagnosis of instructional tasks, monitoring the teaching and learning process) [14]. Additionally, results on the efficiency of particular programs are mostly scarce [12]. In order to effectively foster diagnostic competences and adapt university programs, we need to understand the components that constitute diagnostic competences and how these components are interrelated.

1.1. Diagnostic Competences as a Specification of Professional Competence

Until a few years ago, diagnostic competences were often studied in terms of the accuracy of teachers' judgment, with diagnostic accuracy referring to the difference between a teacher's judgment or a teachers' critical observation and more objective assessments of performance [15,16]. However, supporters of a broader understanding of competence criticized that diagnostic competences cannot be limited solely to measures of accuracy (e.g., [15,17]). According to Schrader [14], a broader understanding of diagnostic competence covers the entire process of diagnosing, including the ability to use appropriate strategies and methods of data collection and processing, as well as to interpret the obtained data properly. This understanding of the diagnostic process is also necessary to effectively help prospective teachers develop diagnostic competence [15]. Therefore, judgment accuracy can only be regarded as one component of diagnostic competences within the broader understanding of competence that takes situated contexts into account.

The refinement of additional components of diagnostic competences can be guided by research on the expert paradigm [18]. The shift within teacher professionalism research from a merely cognitive perspective to a situated perspective on professional competence is considered decisive in order to examine teacher competences as closely as possible to the demands of the real world [9,15,19–21]. In an attempt to integrate both perspectives, Blömeke et al. [1] modeled professional competence on a continuum including personal dispositions such as cognition or affect-motivation, that underlie situation-specific skills, which again mediate between teachers' dispositions and their performance. A particular strength is the consideration of those situation-specific skills that teachers require to succeed in specific situations such as diagnosing. This approach offers the advantage of considering diagnostic competences more broadly, rather than limiting investigations to diagnostic abilities [5] or operationalizing diagnostic competences as accuracy of teachers' judgments only [15,16]. Therefore, in accordance with the competence as a continuum model, diagnostic competences can thus be understood as a latent trait including "those dispositions, situation-specific skills and performance that teachers need for diagnosis in the context of teaching and learning" [6] (p. 43).

Following this understanding, several components must be considered when examining diagnostic competences bringing together components studied in research on teachers' professional competence [1] and teachers' judgment accuracy [15,16]. First approaches to define diagnostic competences with regard to the competence as a continuum model mainly referred to teachers' professional knowledge and diagnostic skills such as diagnostic activities [4,6]. In that vein, Heitzmann et al. [4] defined diagnostic competences as "individual dispositions enabling people to apply their knowledge in diagnostic activities according to professional standards to collect and interpret data in order to make decisions of high quality." Therefore, three components of diagnostic competences are highlighted: (1) teachers' cognitive disposition and, in particular, their professional knowledge, (2) the application of diagnostic activities as a specification of situation-specific skills in the context of diagnosing, and (3) the need for a measure of diagnostic accuracy to check the agreement with professional standards (see Figure 1).

In the following sections, we describe the three components of diagnostic competences in more detail before referring to empirical findings concerning the relationship between these components.
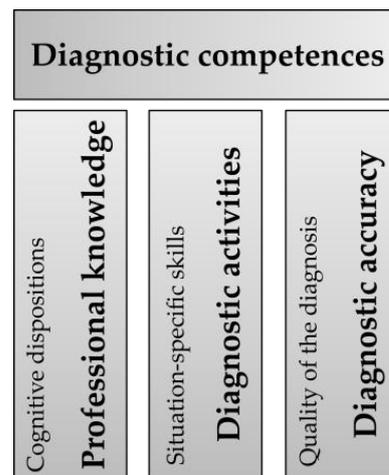
Figure 1. Three components of diagnostic competences, based on [1,4].

1.2. Three Components of Diagnostic Competence

1.2.1. Professional Knowledge

According to Heitzmann et al. [4], knowledge counts as a crucial prerequisite that enables teachers to execute diagnostic activities effectively. Teachers need to apply their knowledge in different diagnostic situations and recall what they know about effective teaching, diagnosing students' (mis)conceptions, and how to support students' learning progress [5,22]. Different facets describe teachers' professional knowledge based on Shulman's division into PK, CK, and PCK [23,24]. The three facets cover knowledge that teachers need for effective teaching. This comprises general pedagogical-psychological knowledge (PK), which is knowledge about classroom management and generic strategies and methods of teaching, learning, and assessment [24–26]; content knowledge (CK) that is knowledge about subject-specific facts, concepts, and methods [27]; and pedagogical content knowledge (PCK) that is knowledge about how to make a particular content accessible for a particular group of students taking into account content-dependent (mis)conceptions of students and instructional strategies of the subject [20,23,28,29]. PK is assumed to be content-independent [30], and therefore, PK seems more relevant for diagnosing general characteristics such as classroom management (cf. [6]). Both CK and PCK are mainly considered subject- and content-specific, and thus, they are applied in subject-specific instructional situations [31,32]. As specific instructional quality features characterize effective science teaching, science teachers' subject-specific PCK counts as the knowledge facet with a high influence on instructional quality and student achievement [33–35]. PCK can therefore be considered a pivotal knowledge facet of teachers' diagnostic competence when the diagnostic focus is on subject-specific instructional aspects. However, regarding the relationship between the three knowledge facets, researchers assume CK to be necessary but not sufficient for the development of PCK, while PK counts as an important precondition to applying CK and PCK in subject-specific instruction [30,31,36]. In order to measure teachers' knowledge facets in a standardized way, different methods such as paper-pencil assessments with multiple-choice or open-ended items, semi-structured interviews, or concept mapping have been used [29,37–39].

1.2.2. Diagnostic Activities

When teachers assess specific instructional situations, they engage in situation-specific diagnostic processes. These processes require the execution of situation-specific skills that have been termed as assessment skills [40], professional vision [41,42] or noticing [43] in the context of classroom assessments, and diagnostic skills [5,6], or diagnostic activities [4,17,44] considering the specific context of diagnosis. Diagnostic activities can be described as those activities teachers execute to evaluate data on, for example, learning conditions and prerequisites of learners in order to optimize the overall instructional pro-

cess [4,44]. With regard to a specific diagnostic situation, different diagnostic activities may be relevant. Diagnostic activities may also vary regarding the weight attributed to each activity and the way these activities are performed [4,45]. This variability of possible activities makes adaptation to specific diagnostic contexts possible. Overall, eight diagnostic activities have been differentiated following scientific reasoning processes: problem identification, questioning, generating hypothesis, constructing artefacts, generating evidence, evaluating evidence, drawing conclusions, and communicating the process/results [4,45]. Descriptions of the eight diagnostic activities can be found in Table 1.

Table 1. Taxonomy of the diagnostic activities according to Heitzmann et al. [4] and Fischer et al. [45]. Note that not each diagnostic activity is appropriate for a given situation, and thus, the number and type of the executed diagnostic activities may vary.

Diagnostic Activity	Description
Identifying problems	A noteworthy event that may influence student learning is noticed by the teacher.
Questioning	The teacher asks questions to find out more about the identified problematic incident or its cause.
Generating hypothesis	The teacher generates a hypothesis about possible sources of the identified problem.
Construct or redesign artefacts	The teacher creates content-specific tasks suitable for identifying underlying instructional problems or detecting students' misconceptions.
Generating evidence	Evidence is generated either by the use of a constructed test or a created task or through systematic observation and description of the problematic incident.
Evaluating evidence	The teacher assesses the generated evidence regarding its support to a claim or theory. He/she interprets the data, thus making sense of the generated evidence with regard to his/her belief, knowledge, and expertise (cf. [46]).
Drawing conclusions	As a result of evaluating evidence, the teacher predicts consequences regarding student learning or makes suggestions for alternative instructional strategies.
Communicating the process/results	The teacher scrutinizes diagnostic results to colleagues, students, or parents.

A more analytical operationalization of situation-specific skills for assessing classroom situations is described in the concepts of "professional vision" and "teacher noticing" (e.g., [41,43]) (see also Section 3.2.2). Overall, situated approaches for the measurement of situation-specific skills such as diagnostic activities that science teachers apply in practices close to the assessment of classroom instruction include, for example, classroom observations, reflection on lesson plans, responding to students' ideas, or video-based analyses [47–49]. All these approaches are based on evidence collected in the specific context in which the skills are applied.

1.2.3. Diagnostic Accuracy

Diagnostic competences are also reflected by the quality of the diagnosis, which can be operationalized by means of accuracy measures [4,17]. Accuracy has often been investigated in terms of judgments accuracy that has been assessed at different levels, for example, at the student-related level by focusing on teachers' judgments about student achievement [15], or at the classroom level by assessing instructional features such as task demands [5]. Considering the student level approach, researchers investigated correlations between teachers' judgments of student characteristics and students' outcomes in a standardized test, or the accuracy of the rank order of student performance according

to competence levels as measures [17,40,50]. At the classroom level, judgments can be compared to specific standards of the domain (e.g., features of instructional quality) in order to consider the quality of the information basis and to obtain a measure of accuracy (cf. [15,51]). However, other measures of accuracy referred to the ability to apply situation-specific skills accurately. For example, researchers focused on perception accuracy, which describes "the precise observation of a professional situation" [2] (p. 373). Carter et al. [52] investigated perception accuracy in terms of immediate perception of science classroom environments of rapidly presented visual classroom stimuli on presentation slides. By limiting the time available for observing the situations, differences in perception accuracy between experts and novices, and thus in the participants' perceptual skill, were revealed. Therefore, accuracy measures should be considered as one component of diagnostic competences as well.

1.3. Empirical Evidence on the Relationships between the Components of Diagnostic Competences

Depending on the diagnostic focus (rather generic: e.g., classroom disruption [53]; or rather subject-specific: e.g., diagnosing biology instruction [54]), either PK or subject-specific facets such as CK or PCK may be relevant to the application of diagnostic activities and diagnostic accuracy (cf. [55]). Tolsdorf and Markic [12] studied different knowledge types (conditional, technological, knowledge of change, competence knowledge) relevant in the context of diagnosing in chemistry. However, they only studied to what extent pre-service teachers that were at different stages of their education differed in, for example, different beliefs, attitudes, and knowledge about the importance of diagnosis in science or knowledge about how to change learning material concerning the needs of the learners. Among others, they found the most positive attitudes toward diagnosis or clear ideas for changing learning materials for more experienced pre-service teachers indicating the role of teacher education in university. However, they did not explicitly investigate other components of diagnostic competence such as diagnostic activities. Furthermore, Tolsdorf and Markic [12] assumed the practical applications of what the students had previously studied and experiences to be crucial for the change of diagnostic competence (defined here rather cognitively) in the course of pre-service teachers' studies. This points to the need to investigate diagnostic competences in situated approaches [21].

Situation-specificity of diagnostic competences was emphasized by Hoth et al. [6]. To investigate teachers' situation-based diagnostic competences, they investigated mathematics teachers' situation-specific skills in order to identify different perspectives of situation-based diagnoses. Furthermore, they examined interrelations between these diagnostic perspectives and teachers' knowledge. Results showed that a content-related perspective in the given classroom situation was related to high average mathematical CK. In addition, teachers using both a didactical and a content-related mathematical perspective had the highest average mathematics PCK scores. Furthermore, there were instances with a higher general PK score associated with a more pedagogical focus on classroom situations, with teachers focusing on aspects of classroom management, organizational aspects, and other pedagogical aspects. Overall, their qualitative analyses suggested that "teachers with greater knowledge not only interpret classroom events more adequately but also knowingly focus their attention to the relevant aspects" [6] (p. 52). Similar results have been found by König et al. [25], who showed that interpreting general pedagogical classroom situations correlated with general pedagogical knowledge. Furthermore, higher values of mathematics teachers PCK were positively connected with noticing relevant teaching and learning incidents, with "relevant" referring to the accurate diagnosing of aspects that matter in terms of students' learning in math [21]. Moreover, teachers with below-average PCK focused on rather superficial characteristics that were irrelevant to the diagnostic problem and thus not accurate. In addition, Blömeke et al. [56] emphasized that pre-service teachers who were prepared to teach both lower- and upper-secondary school (mathematics) had stronger prerequisites and more mathematics-related learning opportunities that resulted in a stronger cognitive base. Furthermore, their cognitive base

(including CK and PCK) was better connected to situation-specific skills. Comparable results on the relationship between knowledge and skills can also be found in research on professional vision (e.g., [38,57]).

Within research focusing on skills relevant during the diagnostic process, Wildgans-Lang et al. [17] showed that the quality of executed diagnostic activities was more important for diagnostic accuracy, and thus, for the quality of the overall diagnosis than the frequency of the diagnostic activities. Furthermore, they differentiated between diagnosing competence levels and diagnosing students' misconceptions and assessed the diagnostic accuracy with regard to both aspects. Results indicated the accurate diagnosis of misconceptions to be more challenging than the accurate diagnosis of competence levels. However, they did not analyze or link knowledge facets in their study but exposed the relation between diagnostic activities and professional knowledge as an important issue for further research.

In educational research, the correlates of diagnostic accuracy such as personal cognitive traits or professional knowledge facets are still not well enough studied [14,58,59]. Accuracy has mostly been considered in terms of comparisons of pre-service teachers' answers/observations with expert ratings/answers that served as a measure of correctness (e.g., [40,55,58]) but without explicitly investigating relationships to other components. In clinical research, content-specific knowledge counts as the basis for diagnosing clinical cases accurately, and accuracy is assumed to depend on skills relevant for correct interpretations [60,61]. Transferred to teacher education, this may imply that subject-specific knowledge facets (CK, PCK) are more relevant for diagnosing subject-specific cases accurately. Accuracy may rely on skills such as the elaborate execution of diagnostic activities that are assumed to be quite poor in pre-service teachers [52,62]. However, so far, only a weak correlation could be found for the relationship between the accurate rank order of tasks and student performance, and teachers' PCK (on text-image integration in biology, geography, and German) [50]. Studies focusing on teachers' adaptive teaching skills in terms of adequate planning and carrying out instruction found moderate correlations with teachers' accuracy of the rank order of student performance [63,64]. While some studies thus focused on student-related accuracy, there is, to our knowledge, a lack of studies on accuracy measures regarding the diagnosis of instructional features in science classrooms that improve the quality of science instruction, and thus, student achievement. Karst et al. [10] traced this back to the greater effort when measuring features of instructional quality that require, for example, the use of videography.

Overall, depending on the diagnostic focus, a correlation between teachers' situation-specific skills, their diagnostic accuracy, and corresponding knowledge facets can be assumed but has not been systematically investigated yet. The use of videotaped classroom situations may be one approach to examine the different components collectively and in a situated way.

1.4. Video-Based Assessment of Diagnostic Competences

For pre-service science teachers (PST) it might be challenging to succeed in diagnostic situations within the complex environment of a science classroom, as PST are less experienced and less skillfully with regard to their situation-specific skills [41,62]. Video-based programs and instruments have been developed, in which videos were used in different ways: to reflect on teachers' own or other teachers' practice; to show best practice training, in which teaching strategies can be observed and adapted for own teaching; or to promote situation-specific skills for the interpretation of important features of classroom interactions [42,65,66]. Researchers assume that supported by the situated context, teachers' professional knowledge can be activated, and necessary situation-specific skills can be applied [2,6,57]. The advantage of videos is that they approximate practice, reduce the complexity of the diagnostic situation, and thus, can promote PSTs' competence development [67]. However, video-based instruments are also considered promising for assessing teachers' diagnostic competences within a situated context, and thus, for measuring their diagnostic activities and diagnostic accuracy [2,6]. Investigation and training

of diagnostic competences within learning environments have often been student- or interaction-based [68]. Diagnostic tasks within video-based instruments referred mostly to the diagnosis of teacher-student interaction on mathematical content (e.g., [6,40,57]), or students' thinking (e.g., [37]), but diagnostic contexts regarding the assessment of the instructional behavior of the teacher with regard to features indicating instructional quality that impact learning have rarely been applied. Researchers, however, emphasized the relation between teacher knowledge and instruction. Knowing about effective instructional strategies and being able to diagnose instructional situations and offer effective instructional alternatives is considered a crucial skill that was found a significant predictor of student learning [57,69]. A first approach can be found by Meschede et al. [38], who assessed pre-service science teachers' situation-specific skill professional vision with regard to the two instructional aspects cognitive activation and structuring learning situations. However, the authors claimed that future research needs to conceptualize and investigate assessment skills with respect to other aspects of instructional quality or in other content-specific domains. The study of diagnostic competences and its components with respect to further subject-specific features of instructional quality is still pending.

1.5. Summary

Overall, three observations can be made: First, investigating diagnostic competences should go beyond judgment accuracy and should also include components such as professional knowledge and diagnostic activities to fully grasp the construct. Second, previous analyses mostly referred to individual components of diagnostic competences, for example, teachers' situation-specific skills for making situation-based diagnoses [6]. Third, considering all components and their relationships is helpful to plan further studies in the future by building on existing relations. This aspect can be considered particularly worthwhile in relation to teacher education, in which diverse offers and courses for the development of competence are embedded without systematically addressing (aspects of) diagnostic competences [13]. Knowing about components that influence teachers' diagnostic competences and which can be modified in teacher education is therefore of great importance [70].

2. Aims and Hypotheses

As described in the previous sections, effective diagnoses, and thus, diagnostic competences, are considered to depend on the activation of appropriate knowledge facets that underly the execution of diagnostic activities (cf. [6,38]). Since concrete empirical evidence regarding diagnostic activities, the inclusion of accuracy measures in situated approaches, as well as interrelations between the professional knowledge base, diagnostic activities, and diagnostic accuracy as components of diagnostic competences is still rare, we want to address these issues within a biological context in higher education since the development of competence needs to start within university teacher education [12,13]. Thus, the present study makes an effort to measure the three components of pre-service biology teachers' diagnostic competences in order to investigate the relationship between them as starting point for a programmatic approach to the investigation of pre-service biology teachers' diagnostic competences (cf. [71]). Therefore, we addressed the following research question:

How do the different knowledge facets PCK, CK, and PK relate to diagnostic activities and diagnostic accuracy?

Considering the demand for practical approaches within the investigation of PST diagnostic competences [6,40], we investigated this question within a situated context using videotaped classroom situations showing whole-class biology instruction. A whole-class diagnostic focus is considered more complicated than diagnosing a particular student or a group of students [12] but reflects teachers' everyday-life practice. Therefore, it is necessary to understand how components of diagnostic competences relate to each other in order to develop or adapt university programs. The following hypotheses were derived from previous research:

Hypothesis 1. *With regard to the assessment of science instruction (cf. [38]) and previous findings on the importance of content-related knowledge facets for subject-specific instructional quality [33,34], PCK can be considered as a pivotal knowledge facet of teachers' diagnostic competences when the diagnostic focus is on subject-specific instructional aspects. Therefore, we assume PCK to be strongly related to the application of diagnostic activities and diagnostic accuracy [6,21,35,38,55,57].*

Hypothesis 2. *CK was found to be a necessary condition for PCK development, but research about its relation to diagnostic activities or accuracy is scarce. Therefore, we assume CK to be correlated with PCK, while a connecion between CK and diagnostic activities or diagnostic accuracy is not assumed (cf. [30,33,34,36]).*

Hypothesis 3. *Finally, a relationship between PK, diagnostic activities, and diagnostic accuracy should not exist, since the focus of the diagnosis lies on subject-specific instructional quality, and thus, not on general aspects of teaching and learning (cf. [6,41]).*

3. Materials and Methods

3.1. Design and Sample

Data collection was embedded in a mandatory seminar attended by pre-service biology teachers at the beginning of their teacher education. Using the video-based assessment tool DiKoBi *Assess* (German acronym for diagnostic competences of biology teachers in biology classrooms) was compulsory for all PST. Still, participation in the study, and thus, releasing their data for analysis was voluntary. All participants signed informed consent documents stating an anonymous and voluntary participation.

The present study had a cross-sectional design with two points of data collection. According to Spector [71], cross-sectional designs are most useful to provide initial evidence of the extent to which variables "are related without introducing the complexities of temporal flows that might distort relationships" (p. 130). Since we were interested in understanding the relationships between the three components of diagnostic competences which we defined, the cross-sectional design provided a useful starting point within our research that will become more complex in subsequent studies.

First, we asked PST of the subject biology to complete three professional knowledge tests to measure PCK, CK, and PK. Second, we used the video-based assessment tool DiKoBi *Assess* to measure PSTs' diagnostic activities and diagnostic accuracy. This means the data set allowed the computation of five different PST measures (a PCK measure, a CK measure, a PK measure, a diagnostic activities measure, a diagnostic accuracy measure).

In DiKoBi *Assess*, PST had to use diagnostic activities to diagnose a biology teacher's subject-specific instruction within a real-life teaching situation to capture diagnostic competences as ecologically valid as possible (cf. [72]). The sample consisted of 186 PST of the subject biology (72.0% female; average study semester: $M = 3.3$, $SD = 1.3$; age in years: $M = 23.0$, $SD = 3.8$). A percentage of 36.6% of the PST strove for a certification qualifying them for the academic track of German secondary education (German Gymnasium), and 63.4% attended programs for the non-academic track that qualifies students for a vocational career. For an overview of the German school system, see Cortina and Thames [73].

3.2. Measurements

3.2.1. Professional Knowledge Tests

Professional knowledge was assessed through the use of paper-pencil tests that included open-ended items (responses were written in text fields), single best answer (SBA) items: one correct answer must be selected from a set of possible responses consisting of multiple distractors and one correct answer), or multiple true/false items (all of the possible responses must be assessed for their validity) [74].

The PCK- and CK-tests considered the biology topic *skin* as this was the same topic covered in the video-based assessment tool DiKoBi *Assess*. The tests were adapted versions

of the professional knowledge tests, which have been utilized in the ProwiN project (Professional Knowledge of Teachers in Science) [75,76]. In the tests, aspects of PSTs' declarative and action-related knowledge were assessed. Declarative knowledge (knowing *that*) includes knowledge about terms, facts, and principles (e.g., listing the advantages and disadvantages of a specific model); action-related knowledge (knowing *how*, knowing *when and why*) is needed for successful instruction in different situations [77]. Knowing *how* is knowledge about an individual science teacher's (instructional) practices and processes (e.g., knowing how to deal with student ideas). Knowing *when and why* refers to "knowledge about conditions under which decisions and practices are appropriate and knowledge about reasons for performing specific practices" [77] (p. 6).

The PCK-test covered knowledge of instructional strategies and knowledge of student (mis)conceptions. Both issues count as important components of teachers' PCK [22,31]. Utilizing the model of Tepner et al. [78], eight open-ended items and five SBA items concerning three PCK dimensions were included in the test (see Table 2). The CK-test included 13 open-ended items and 15 SBA items. Topics that were covered with the items are shown in Table 3. Criteria for item scoring of both the PCK- and CK-test were provided in two separate coding manuals. Two independent raters used the coding manuals to code ten percent of both the PCK- and CK-tests. Results of two-way random intra-class correlations ($ICC_{absolute}$) showed a high agreement between the two raters (PCK-test: $ICC_{absolute}(310,310) = 0.84$, $p < 0.001$; CK-test: $ICC_{absolute}(341,341) = 0.97$, $p < 0.001$) [79].

Table 2. Overview about the topics covered in the Pedagogical Content Knowledge (PCK)-Test (based on [75,76,78]).

Item Label	Topic	PCK-Dimension
PCK-1a °	Causes of a specific student error	
PCK-1b °	Dealing with specific student errors	Student errors
PCK-1c °	Further topic-specific student errors	
PCK-2a °	Advantages and disadvantages of a specific model on the structure of the skin	
PCK-2b °	Critical reflection of the model	Use of models
PCK-2c °	Setting goals for elaborated use of models	
PCK-3a *	Assessment of an experiment (cold protection through fat) regarding subject-specific aspects (e.g., steps of scientific inquiry)	
PCK-3b °	Alternative instructional implementation of the experiment	Use of experiments
PCK-3c °	Further student experiments on the topic *skin*	

° Item was an open-ended item. * Item label was further differentiated into five single best answer (SBA) items.

Table 3. Overview about the topics covered in the Content Knowledge (CK)-Test (based on [76]).

Item Label	Topic
CK-1 *	Control and regulation of skin functions
CK-2 °	Regulation of body temperature
CK-3 °	Structure and function I
CK-4 *	Structure and function II
CK-5 *	Human skin color
CK-6 °	Transfer of biological concepts/principles

* Item label was further differentiated into five SBA items. ° Item label included open-ended items (one for CK-2; nine for CK-3; three for CK-6).

The knowledge facet PK was assessed by utilizing a paper-pencil test that was adapted from the BilWiss project [80,81]. The adapted version contained one out of six different dimensions originally used in BilWiss. For this study, the short scale of the dimension *instruction* was used, because it contained items about generic features of instructional quality such as classroom management, supportive climate, and cognitive activation, which are referred to as basic dimensions of instructional quality [33,82,83]. The instrument also contained items concerning general pedagogical issues of teaching such as teaching methods (see Table 4). Therefore, the selected dimension of the BilWiss test was the most important one with regard to the differentiation between generic and subject-specific features of instructional quality, which was important for accurate diagnosis of the videotaped classroom situations. The PK-test contained five SBA items and ten multiple true/false items. The item scoring followed the instructions from BilWiss [80]. Data sets from the BilWiss project, in which the PK test from this study was developed and used, are publicly available on the IQB website [81].

Table 4. Overview about the utilized items from the Pedagogical Knowledge (PK)-Test (based on the short scale of the dimension *instruction* from the BilWiss project [80]).

Item Label	Topic	Item Label	Topic
PK-1 *	Structuring feedback to students	PK-9 *	Supportive climate/motivating learning environment
PK-2 °	Discourse management and question formats	PK-10 *	Participation of the students (working with weekly schedules)
PK-3 °	Methodological organization, social forms, research on teaching methods	PK-11 *	Support of learning processes (role of students' prior knowledge)
PK-4 *	Preparation of variable/differentiated teaching progressions	PK-12 *	Cognitive activating learning goals according to Bloom
PK-5 *	Methodological forms of teaching (advantages of project work)	PK-13 *	Transparency of goals and requirements
PK-6 °	General didactics: constructivist models	PK-14 °	Models of teacher-student interaction
PK-7 *	General didactics: theoretical models of teaching and learning	PK-15 °	Classroom management
PK-8 *	Constructive handling of errors/analysis of student errors		

° Item was a multiple true/false item. * Item was an SBA item.

The three tests were evaluated utilizing Rasch theory and Rasch analysis techniques [84,85]. The Rasch Partial Credit Model (PCM) was used utilizing the Winsteps program [86]. The use of Rasch allowed "person measures" to be computed for each respondent for each instrument. Therefore, the data collected from the PCK-, CK-, and PK-tests allowed PCK, CK, and PK Rasch measures to be computed for each respondent. Reasons for utilizing Rasch are that raw scores (be it from a multiple-choice test, a partial credit test, or a rating scale) are non-linear. Rasch allows one to take that non-linear data and compute Rasch person measures which are expressed on a linear scale [85,87]. It is those linear measures that are needed for parametric statistics. Additional reasons why Rasch analysis techniques should be used when test and survey data is evaluated include that Rasch methods (1) express items and respondents on the same measurement scale, (2) provide wide-ranging Rasch indices to evaluate the functioning of items, (3) allow the computation of the measurement error of each item and each respondent, (4) allow respondent measures to be computed even if data is missing, (5) correct for the non-linearity of raw test scores, and (6) enable alternate forms of an instrument to be developed (through linking items), such alternate forms enable respondent performance to be expressed on the same scale regardless of test form completed [85,87].

One important component of an analysis utilizing Rasch methods is an assessment of the "fit" of items. To evaluate data fit, item Outfit-MNSQs (mean-squares) were utilized. Additionally, Rasch person reliability and Rasch item reliability were computed and evaluated. It has been argued that for a productive measurement, item Outfit-MNSQ values should not exceed 1.5 [88]. High values of item reliabilities demonstrate that both the range of item difficulty and the sample size are appropriate to measure the variables precisely. Person reliability is impacted by the length of the test and the range of abilities of respondents [86]. Item fit statistics of the knowledge tests showed good fit values in which all items exhibited an Outfit-MNSQ below 1.5 (PCK: 13 item outfit-MNSQ < 1.18; item reliability = 0.96; person reliability = 0.55; CK: 28 item outfit-MNSQ < 1.35; item reliability = 0.97; person reliability = 0.67; PK: 15 item outfit-MNSQ < 1.34; item reliability = 0.98; person reliability = 0.50).

3.2.2. Video-Based Assessment Tool DiKoBi Assess
Measuring Diagnostic Activities

To measure PSTs' diagnostic activities, we used the video-based assessment tool DiKoBi *Assess* that was embedded in an online-survey platform [89]. DiKoBi was developed to provide diagnostic situations of real-world demands for biology teachers, in which subject-specific knowledge and skills can be applied to assess biology teaching [54]. Six staged videos were embedded in DiKoBi to address six different challenging situations that biology teachers have to confront when teaching biology (e.g., elaborate use of three-dimensional models). All of these six challenging situations addressed the biological topic "skin," each challenge concerning a different subject-specific dimension of instructional quality. The embedded dimensions and including features have been found important factors that impact student achievement [90]. These dimensions are

(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.*

In DiKoBi, every challenging situation to be diagnosed started with the video of the classroom situation. Afterward, PSTs had to complete three tasks. Each task required the execution of a diagnostic activity. Since not all diagnostic activities were found to be useful for diagnosing in DiKoBi, a selection of three activities that can be considered crucial in the context of video analyses was made in comparison with conceptualizations reported in research previously [91]. Besides conceptualizing situation-specific skills as perception, interpretation, and decision-making, originated from the competence as a continuum model [1], other conceptualizations exist. These refer to reflective skills for video viewing or classroom observation and are discussed, for example, under the term professional vision, which includes several critical activities [42,92]. Within a teaching context, professional vision means the ability to *notice* (that is paying attention to relevant events in the classroom) and to *reason* about relevant features of classroom interaction [41]. The reasoning process is knowledge-based and can be differentiated into three further activities: *describing* the situation without making judgments, *explaining* the situation by linking the observation to professional terms and concepts, and *predicting* possible consequences from the observed situation [41,43]. Considering the skills displayed in the competence as a continuum model and the aforementioned activities described within the concept of professional vision, four diagnostic activities (DA) can be considered crucial within diagnostic processes. First, science teachers have to identify noteworthy events in the science classroom (DA = *problem identification*), and systematically observe and describe the noteworthy events (DA = *evidence generation*). Second, the teachers explain the situation by actively drawing on their declarative and action-related knowledge (e.g., by using professional terms and theories of teaching and learning that support the relevance of their

observation) (DA = *evidence evaluation*). Third, they make decisions on how to continue with instruction or respond to students' activities, or they even have to propose alternative teaching strategies (DA = *drawing conclusions*) (cf. [2,93]). Since problem identification occurs rather invisibly in the participant's mind, we assumed problem identification to be indirectly included in the events described without explicitly measuring it. This assumption is based on the fact that in order to describe an incident, a teacher must show awareness of exactly this incident or problem that occurred [92].

Three tasks prompted the diagnostic process in DiKoBi. Each task focused on a specific diagnostic activity. For Task *Describe* (DA = *evidence generation*), the PST had to identify and describe challenging aspects of each classroom situation; for Task *Explain* (DA = *evidence evaluation*, the PST had to reason about their described challenges by linking their description to scientific theories and concepts; and for Task *Alternative Strategy* (DA = *drawing conclusions*), the PST had to propose an alternative teaching strategy, and give reasons why this would improve instruction. For a more detailed description of the development and the design of DiKoBi, see Kramer et al. [54].

A coding manual was utilized to analyze the PSTs' answers to each task of the six challenging situations. The coding manual is based on subject-specific instructional quality features and corresponding indicators described in the science literature (e.g., using challenging tasks to foster conceptual understanding [83]). In empirical studies, a positive impact of these features on student learning has been found. Examples from the manual and corresponding references can be found in Kramer et al. [54]. For each task, the answers were coded according to the content-related knowledge facets (PCK, CK, PK), and the quality of the appropriate diagnostic activity was assessed. Zero (0) points were used to indicate very low (not accurate) answers. For correct answers of improved quality, 1 or 2 or 3 points could be utilized for a rating. Appendix A provides an overview of the procedure utilized for coding and the codes of the quality levels that have been used for coding the PSTs' answers.

Three independent raters used the coding manual to code statements of Task *Describe*, Task *Explain*, and Task *Alternative Strategy* to ensure objective coding of the answers. 337 statements from 10 PST were coded by all three raters. Results of a two-way random intra-class correlation ($ICC_{absolute}$) analysis of these ratings suggested a high agreement between the three raters ($ICC_{absolute}$ = 0.90, F (1520, 3040) = 10.26, $p < 0.001$, N = 1521) [79]. For the small number of discrepancies in coding that were observed, these differences were discussed by all three raters prior to the rating of the remaining data by a single coder. Complex cases continued to be discussed together during the ongoing coding process.

After coding the different tasks and situations, the research team calculated Rasch person ability measures for each respondent. Similar Rasch techniques as that described previously for the PCK-, CK-, and PK-tests were utilized. This Rasch person measure that we computed expressed the level of each PST's ability to execute diagnostic activities accurately. Thus, the Rasch person measure provides an assessment of the PSTs' diagnostic level utilizing the data collected for *evidence generation*, *evidence evaluation*, and *drawing conclusions* for each of the six classroom situations. Fit statistics of the Rasch model showed productive measures (diagnostic activities: 29 item outfit-MNSQ < 1.43; item reliability = 0.95; person reliability = 0.76).

Calculating Diagnostic Accuracy

On the one hand, accuracy was already taken into account in the scoring of the diagnostic activities. Thus, it was included as a quality criterion in the measurement of the diagnostic activities, since zero points were assigned if the description of the pre-service teachers referred to a not-accurate observation. However, for investigating the relationship between components of diagnostic competences, we established another, more explicit measure of accuracy, which is operationalized as perception accuracy, referring to the precise perception and description of biology instruction (cf. [52,94]). Therefore, diagnostic accuracy was assessed by a teachers' individual observation compared to the

scripted challenges embedded in the videos. These scripted challenges and corresponding indicators of instructional quality features served as an objective criterion for comparison since they were derived based on empirically proven features for effective teaching. In other words, we examined whether PSTs' descriptions of the challenging situations referred to the PCK-challenges that addressed missing features of subject-specific instructional quality. Descriptions that referred to superficial or general pedagogical observations not relevant for teaching and learning in the specific situation were not counted. To better understand the accuracy calculation, we will illustrate the procedure with an example: Given the classroom situation *use of models*, the two challenging instructional aspects *elaborate model use* and *critical reflection* were included as subject-specific features of instructional quality. Indicators addressing the lack of an *elaborated model use* were statements such as "The model is used for illustrative purposes only" or "The model is described incompletely." Indicators addressing the lack of *critical reflection* were statements such as "Teacher does not initiate critical reflection of the model" or "Teacher does not discuss the model with students." PST could describe up to ten observations per classroom situation. Subsequently, each described observation was compared with indicators of subject-specific instructional quality listed in the manual. If this was the case, the response was scored with one point. If this was not the case, zero points were given. Subsequently, all points (i.e., correct observations) for a challenging classroom situation such as *use of models* were added up and divided by the overall number of observations made in this classroom situation. For example, a participant who described four observations in the classroom situation *use of models*, but only two of them were indicators of biology-specific instructional quality and thus considered accurate, the calculated diagnostic accuracy of this classroom situation was 0.5. In the end, we calculated the average of the accuracy measures of the six classroom situations for the final measure of diagnostic accuracy.

3.3. Data Analysis

For the knowledge variables PCK, CK, and PK, as well as for the variable diagnostic activities, we conducted Rasch analyses [84] using the software Winsteps 3.81 [86]. As mentioned, we computed Rasch person measures, which we utilized for our subsequent statistical analysis. To test our hypotheses, we used path analyses in AMOS 26 [95] with the equal interval person abilities resulting from Rasch analysis of PCK, CK, and PK as predictor variables and the person abilities of diagnostic activities as well as the calculation of diagnostic accuracy as outcome variables. The model was estimated with maximum likelihood. For model fit, we used the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). Model fit was estimated by guidelines of Hu and Bentler [96]: CFI > 0.90, RMSEA < 0.05, SRMS < 0.08. Results of the path analyses are shown as standardized values.

4. Results

An overview of all variables of the path model, including means and standard deviations, and correlations are shown in Table 5. For the variables PCK, CK, PK, and diagnostic activities, we used person abilities from the PCM that take item difficulties into account [85]. Negative mean values for PCK, CK, and diagnostic activities indicate that the tests used were rather difficult for the PST in our sample. There was a moderate correlation between PCK and almost all other variables, indicating the great importance of PCK for diagnostic competences. Additionally, there was a strong correlation between diagnostic activities and diagnostic accuracy. This observation can partly be explained by the fact that accuracy was also considered when scoring the executed diagnostic activities (see Appendix A). For example, if a PST described an incident that has been found as non-problematic according to the coding manual, the PST's description was assessed as "not accurate." Additionally, there were small correlations of CK and PK with diagnostic activities and diagnostic accuracy.

Table 5. Mean scores, standard deviations, and correlations of all path-model variables.

	N	M	SD	PCK	CK	PK	Diagnostic Activities	Diagnostic Accuracy
PCK [a]	186	−1.21	0.72	-				
CK [a]	186	−0.88	0.71	0.30 **	-			
PK [a]	186	0.47	0.50	0.29 **	0.15 *	-		
Diagnostic activities [a]	186	−1.53	0.67	0.36 **	0.22 **	0.25 **	-	
Diagnostic accuracy	186	0.28	0.15	0.30 **	0.22 **	0.21 **	0.70 **	-

[a] Person abilities of variables scaled according to the PCM in Logits. * Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

To investigate the relationship between the professional knowledge facets, diagnostic activities, and diagnostic accuracy, we calculated a path model. Figure 1 shows the model with standardized parameter estimates and levels of significance. The model demonstrated that PCK (β = 0.29, *SE* = 0.07, *p* < 0.001) and PK (β = 0.15, *SE* = 0.09, *p* = 0.031) was significantly related to PSTs' diagnostic activities; both PCK and PK together explained 17% of the variance of diagnostic activities (R^2 = 0.17). Furthermore, 12% of the variance of diagnostic accuracy was attributable to PCK (β = 0.23, *SE* = 0.02, *p* = 0.002). The model had no degrees of freedom, and its fit values were CFI = 1.000, RMSEA = 0.315, SRMR = 0.000. The rejection of the model by RMSEA could be due to the sample size smaller than *N* = 250 [96].

However, whereas results confirmed hypotheses 1 and 2, hypothesis 3 must partly be rejected since the predictor variable PK was positively related to diagnostic activities as well (see Figure 2).
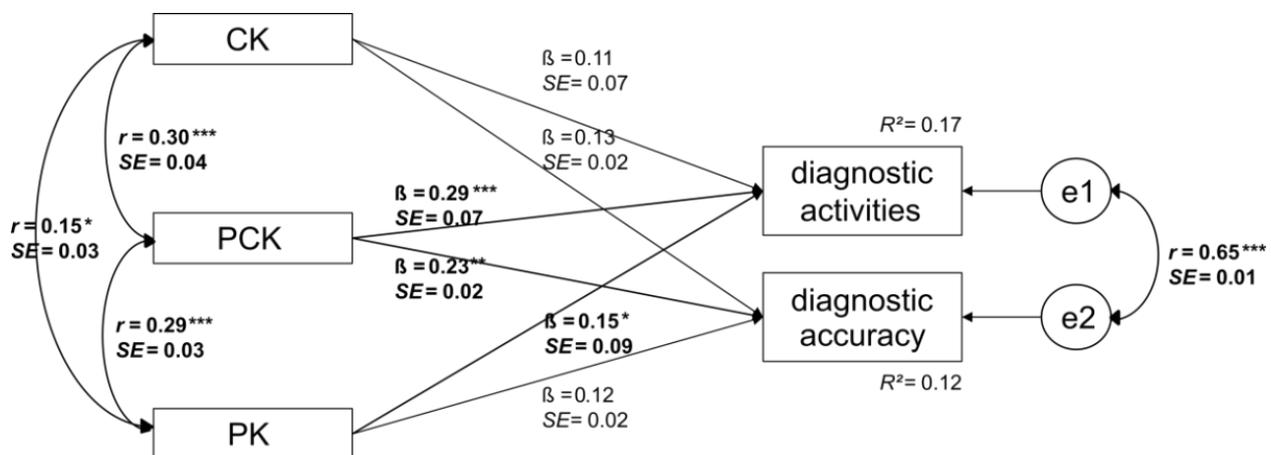
Figure 2. Path model showing a positive relationship between PCK and diagnostic activities (*** *p* < 0.001), between PCK and diagnostic accuracy (** *p* < 0.01), and between PK and diagnostic activities (* *p* < 0.05).

5. Discussion

This study aimed to contribute to teacher education by investigating the relationship between three components of diagnostic competences that have been defined as professional knowledge (PCK, CK, PK), diagnostic activities, and diagnostic accuracy [4]. Following the understanding of competence as described in the competence as a continuum model, we assumed professional knowledge as part of teachers' cognitive dispositions related to teachers' application of diagnostic activities and their diagnostic accuracy [1,6]. However, since we chose a whole-class diagnostic focus on biology instruction in our study, not all knowledge facets were assumed to be equally related to diagnostic activities and diagnostic accuracy. Therefore, PCK, CK, and PK of pre-service biology teachers (PST) were measured with adapted versions of objective, reliable, and valid paper-pencil

tests [76,80]. Diagnostic activities as an operationalization of situation-specific skills were measured with DiKoBi *Assess*. This video-based assessment tool provided subject-specific challenges that a biology teacher has to deal with in real-life instruction. The challenges focused on empirically proven features of instructional quality in the science classroom. As part of their diagnostic competences, pre-service teachers need to know about effective instructional strategies, but they also need to be able to diagnose instructional situations and offer effective instructional alternatives. Being able to apply skills such as diagnostic activities was found a significant predictor of student learning [57,69]. With regard to the aspects of professional vision (description, explanation, prediction), and with regard to the situation-specific skills perception, interpretation, and decision-making that are depicted in the competence as a continuum model, tasks in the video-based assessment tool prompted the use of the diagnostic activities *evidence generation*, *evidence evaluation*, and *drawing conclusions* as relevant situation-specific skills in the context of video analysis and diagnosis of classroom instruction [2,42]. Following this, we want to discuss the results of our path analysis.

Results showed that PSTs' PCK was positively related to the application of diagnostic activities (hypothesis 1), thus further supporting the tendency in existing results on the relation between PCK and situation-specific skills in domain-specific situations [22,38,57]. Furthermore, our results also highlight the importance of PCK for diagnostic accuracy (defined as precise perception and description of incidents relevant in biology instruction). Similar findings have been reported by Hoth et al. [21], who already indicated a connection between high subject-specific knowledge of teachers and their ability to reason about student errors more accurately, "while teachers with low knowledge focus on aspects that are not directly connected to the student's learning" [21] (p. 1). However, it must be noted that both PCK and CK were included in Hoth et al.'s knowledge conception. Therefore, our results explicitly indicate the relationship between PCK and diagnostic accuracy, not only for student errors but also for other dimensions of subject-specific instructional quality. Since many incidents occur simultaneously or in very short succession in biology classrooms, being able to focus on relevant incidents is crucial in terms of a biology teachers' effective instructional behavior [13]. Pre-service teachers' diagnostic accuracy might therefore also be considered important for implementing instructional quality in real-life teaching that should be studied in future research.

The results also showed that for the application of diagnostic activities and in terms of diagnostic accuracy, CK is not critical, but CK is moderately connected to PCK (Hypothesis 2). This result also confirms previous research findings highlighting the role of CK in defining the scope of the development of PCK [30,31,34].

Contrary to our hypothesis 3, the knowledge facet PK was positively related to the application of diagnostic activities. To understand this relation, we want to take a closer look at the utilized items (see Table 2; Table 4). Items of the PCK test referred to three PCK dimensions, which were *use of models*, *use of experiments*, and *student errors* [78]. The three PCK dimensions were covered in the video-based assessment tool as well. However, the video-based assessment tool also covered additional dimensions of PCK, which were *level of students' cognitive activities and creation of situational interest*, *use of technical language*, and *conceptual instruction*. An extension of the PCK paper-pencil test would be useful to cover the same PCK dimensions in both measurement instruments. Subsequently, it would have to be verified whether the present relationships remain the same. On the other hand, two items of the PK test showed similarities to PCK-dimensions utilized in the video-based assessment tool. For example, item PK-11 referred to the role of activating students' prior knowledge for instruction. Activating students' prior knowledge can be considered important in terms of *cognitive activation*, which is one of the three basic dimensions of instructional quality that have been described for effective instruction relevant in different domains [33,83]. At the same time, the implementation of *cognitive activation* has to be concretized from a subject- and content-related perspective [11,90]. Such content-related concretization was done in the corresponding classroom situations

of the video-based assessment tool that referred to the *level of students' cognitive activities* and *conceptual instruction*. That is why those two dimensions were assigned to PCK. In the same vein, there is an overlap between PK-8 (*constructive handling of errors/analysis of student errors*) and the PCK dimension *dealing with (specific) student ideas and errors*. Again, this is a content-related concretization with regard to very specific student ideas and errors on the topic *skin*. Thus, whereas in PK-8 the general principle of dealing with errors was focused, in the items PCK-1a, PCK-1b, and PCK-1c it was about the application of the principle in a subject-specific context. Therefore, there is some overlap in the operationalization of PK and PCK at this point that may be the reason for the unexpected relationship between PK and diagnostic activities.

However, the results might also be interpreted in the sense that both PK and PCK are positively related to the application of diagnostic activities. Considering the situated diagnostic context that we used to assess the diagnostic activities, the multidimensional nature of the teachers' classroom performance might have activated both scientific and pedagogical concepts as suggested by Depaepe et al. [20]. We believe PK is mainly correlated with *evidence generation*, whereas for *evidence evaluation* PCK should be decisive (cf. [25]). Therefore, coding the diagnostic activities as a total person ability measure, as done in this study, may have influenced the result as well. A next interesting step would be to investigate the relation between teachers' knowledge facets and the different diagnostic activities as it can be assumed that subject-specific knowledge is more important for *evidence evaluation* than for the other diagnostic activities (cf. [25]).

A well-developed PK including knowledge of general principles of teaching and learning may thus be seen as a precondition for being able to apply knowledge of content-related basic dimensions to the diagnosis of subject-specific situations (cf. [11,30]). Here, further research can follow that examines the use of knowledge facets in relation to different dimensions of instructional quality in which diagnostic activities are applied. Content-related dimensions such as *cognitive activation* could be related to both PCK and PK, while subject-specific dimensions such as *use of models* or *use of experiments* could correlate more strongly with PCK (cf. [11]). Since the video-based assessment tool DiKoBi *Assess* provided diagnostic situations in the biology classroom, the higher effect sizes for PCK (see Figure 2) and the moderate correlations between PCK and the other components of diagnostic competences indicated the importance of subject-specific knowledge for effective diagnosing in the science classroom.

As is the case in all studies, there are of course some limitations to our study. First, DiKoBi *Assess* was used to measure the diagnostic competences of a sample of PSTs within a situated context. Since the videos were scripted and presented specific challenges in a very condensed way, there might be a gap between the classroom situations in the videos and real-life teaching, even when teachers perceived the classroom situations as authentic [54]. By using short videos of classroom situations, the complexity of diagnostic situations can be reduced. In our case, this was done by focusing on specific challenges of instruction within one video, so that the complexity of each diagnostic situation was reduced by "breaking down practice into its constituent parts for the purposes of teaching and learning" [67] (p. 2058). This procedure may be beneficial for inexperienced PSTs' learning [40], but it must be kept in mind that instruction within the real classroom may involve solving more complex challenges.

Furthermore, the described relationships apply to the selected diagnostic activities that were relevant for diagnosing the classroom situations embedded in the video-based tool. Prompting other activities might change results. In future studies, we also want to include affective-motivational variables such as teachers' beliefs, motivation, and self-related cognitions that may impact teachers' diagnostic activities as well [7,38,94].

Another potential limiting factor to our study could be the person reliabilities of our knowledge tests. Many factors can impact Rasch person reliability. Often such reliability values are impacted by the targeting of a test. If a test is too difficult, or too easy for a sample, the person reliability can be impacted. Such off-targeting of a test to a sample's

ability level is not uncommon in studies. A common rule of thumb is that there should be less than a 1 logit difference between the average person measure on a test and the average item measure [97]. For our instruments, only the value of the PCK-test violated this rule of thumb. The difference between the average person measure and the average item measure was −1.22, meaning that the test was too hard for the respondents. Since the PSTs of this study were at the beginning of their studies, higher reliabilities might be observed if the test instruments were administered later in the curriculum of the PSTs. Linacre [86] has also suggested that an instrument's reliability might be increased if longer versions of a test are utilized, or if a sample with greater variance in ability took an instrument. Since our sample's ability range was rather narrow and PSTs with extremely high or low abilities were not part of the sample, the measure of person reliability decreased.

Furthermore, we did not investigate what pre-service teachers are able to implement in real-life performance. Thus, being able to diagnose instructional features of biology instruction does not necessarily mean being able to implement those features in real-life instruction. The linking of diagnostic competences and practical implementation as well as the investigation of the effects of diagnostic accuracy on instructional quality can follow such considerations (cf. [14]).

6. Implications and Further Research

The results serve as a basis for further, more complex investigations of science teachers' diagnostic competences within the COSIMA *project* (Facilitating diagnostic competences in simulation-based learning environments in the university context) and can be used to design ways of fostering pre-service teachers' diagnostic competences systematically in the future. Following considerations by Meschede et al. [38], we have investigated diagnosis with regard to further features of instructional quality in the field of biology instruction and thus linked dimensions of subject-specific instructional quality with diagnostic competences. It is now important to investigate in more detail which diagnostic activities are related to which knowledge facet and what differences might exist with respect to different dimensions of instructional quality.

The results of the present study support the importance of PCK for biology teacher education, particularly for pre-service biology teachers' diagnostic activities and diagnostic accuracy. In this regard, video-based tools provide the opportunity to apply knowledge and diagnostic activities to classroom situations, which are particularly relevant for biology instruction. Thus, these tools can be used not only to measure skills (cf. [6,11,38]) but also as an effective way for supporting knowledge acquisition as well as for providing varying opportunities in which diagnostic activities can be applied and trained (cf. [42]). Therefore, besides utilizing DiKoBi as an assessment tool, its potential as a learning tool (DiKoBi *Learn*) will be investigated in future studies. Taking the results of the present study into account, DiKoBi *Learn* has then to provide additional instructional information concerning the subject-specific dimensions of instructional quality that are addressed in the videos used.

At the same time, the present results also point to the relevance of a well-founded pedagogical knowledge base. The findings could serve as an incentive to focus more on general dimensions of instructional quality and activities such as the general description of instruction in the pedagogical training of teachers, while subject-specific features are then given priority in subject-specific courses on the basis of practical examples (such as those provided in the video-based tool DiKoBi). Accordingly, in subject-specific courses, the interpretation and evaluation of subject-specific instruction should gain more weight.

Appendix A

Coding Procedure for Diagnostic Activities

The coding procedure for the individual tasks was as follows:

Task *Describe*, DA = *evidence generation*: (1) For each observed and described challenge, we assessed whether it addressed subject-specific pedagogical content (PCK) or merely pedagogical content (PK). Aspects referring to the subject matter (CK) have not been mentioned. For the assignment to PCK or PK, the descriptions were compared with the itemization (indicators) in the coding manual. Consequently, the descriptions could be assigned either to the scripted, evidence-based PCK challenges which we had recorded and embedded in the video clips [53], or to further PCK-aspects or PK-aspects. (2) We evaluated how well each PST executed the diagnostic activity *evidence generation* to assess the descriptions' quality. The quality of the statements was assessed on three levels (see Table A1).

Task *Explain*, DA = *evidence evaluation*: (1) Depending on the knowledge facet to which the description was assigned to, we evaluated (2) whether the statement was per se an accurate explanation and how well did the given explanation relate to subject-specific pedagogical theories. The quality of the statements was assessed on four levels (see Table A1).

Task *Alternative Strategy*, DA = drawing conclusions: (1) For each described strategy, affiliation to PCK or PK was assessed first by comparing the statements with the itemization in the coding manual. Strategies addressing PCK were further assessed on whether they covered aspects of the scripted PCK-challenges. (2) For quality assessment, we evaluated how well the alternative teaching strategy was set up. The quality of the statements was assessed on three levels (see Table A1).

The code "0" was assigned for not accurate statements. This was the case, for example, when incorrect observations have been made that were not visible in the videos (Task *Describe*), when false or incomprehensible explanations were given (Task *Explain*), or when the described alternative teaching strategy did simply not represent an (appropriate) alternative strategy (Task *Alternative Strategy*).

Table A1. Overview of the quality levels and their operationalizations for the assessment of diagnostic activities.

	Code	Description	Example
	−99	Missing	No answer was given.	-
	0	Not accurate	The given description is not accurate.	Introduction with media.
Task *Describe*	1	Unsystematic, incomplete description	Non-specific, superficial description of the identified event; weak/no focus on specific details	Repetition very short.
	2	Systematic, complete description	Several details of the identified event are described with attributes; a strong focus on specific details (visible in the video)	The students should briefly repeat what was discussed last week, but only superficial, general terms are discussed.

Table A1. *Cont.*

	Code		Description	Example
	−99	Missing	No answer was given.	-
	0	Not accurate	The given explanation is not accurate.	Presence of the teacher is minimized -less activating and motivating -room use is almost not given.
	1	Empty phrase	The statement is more of an everyday phrase than an explanation, partly meaningless.	It doesn't really make sense to say we repeat the last lesson and then stop after two aspects.
Task *Explain*	2	Simple reference to concepts/theories	Appropriate to or based on the corresponding description, the subject-specific pedagogical theory is named as a keyword or embedded as a phrase in a sentence.	The teachers' questions do not allow for cognitive activation.
	3	Comprehensive explanation	Observation and theory are related to each other.	Calling one student is not enough, The teacher neither asked for explanations nor did she engage the students to recognize or call on conceptual connections. The activation of prior knowledge could be extended to activate the students more deeply.
	−99	Missing	No answer was given.	-
	0	Not accurate	An alternative strategy is missing or the statement contains a suggested strategy that does not make sense with regard to the context.	To handle the whiteboard; best if the room is empty try it out and see how it moves.
Task *Alternative Strategy*	1	Just one alternative (described non-specific)	One alternative strategy is described. The description is rather general. Concrete examples/references are missing.	Introducing the lesson with a close-up of the skin that arouses interest.
	2	Several alternatives (described in detail and with examples)	Several appropriate suggestions for improvement are described with concrete examples and references; it is explained to what extent the alternative strategy improves what has been criticized before.	Promoting the motivation of the students by using different skin types (elephant skin, crocodile skin), activating prior knowledge (e.g., describing similarities/differences of the different skin types), asking for students' prior experiences (e.g., experiments on the sense of feeling/touch).

References

1. Blömeke, S.; Kaiser, G.; Clarke, D. Preface for the Special Issue on "Video-Based Research on Teacher Expertise". *Int. J. Sci. Math. Educ.* **2015**, *13*, 257–266. [CrossRef]
2. 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]
3. Klug, J.; Bruder, S.; Kelava, A.; Spiel, C.; Schmitz, B. Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teach. Teach. Educ.* **2013**, *30*, 38–46. [CrossRef]
4. Heitzmann, N.; Seidel, T.; Opitz, A.; Hetmanek, A.; Wecker, C.; Fischer, M.R.; Ufer, S.; Schmidmaier, R.; Neuhaus, B.J.; Siebeck, M.; et al. Facilitating Diagnostic Competences in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education. *Frontline Learn. Res.* **2019**, *7*, 1–24. [CrossRef]
5. Brunner, M.; Anders, Y.; Hachfeld, A.; Krauss, S. The Diagnostic Skills of Mathematics Teachers. 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. 229–248, ISBN 978-1-4614-5148-8.
6. 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]
7. Klug, J.; Bruder, S.; Schmitz, B. Which variables predict teachers diagnostic competence when diagnosing students' learning behavior at different stages of a teacher's career? *Teach. Teach.* **2016**, *22*, 461–484. [CrossRef]
8. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK). Ländergemeinsame Inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung [Common State Requirements for the Professional Disciplines and Subject Didactics in Teacher Education]. 2008. Available online: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf (accessed on 23 February 2021).

9. Herppich, S.; Praetorius, A.-K.; Förster, N.; Glogger-Frey, I.; Karst, K.; Leutner, D.; Behrmann, L.; Böhmer, M.; Ufer, S.; Klug, J.; et al. Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teach. Teach. Educ.* **2018**, *76*, 181–193. [CrossRef]

10. Karst, K.; Schoreit, E.; Lipowsky, F. Diagnostische Kompetenzen von Mathematiklehrern und ihr Vorhersagewert für die Lernentwicklung von Grundschulkindern [Diagnostic Competencies of Math Teachers and Their Impact on Learning Development of Elementary School Children]. *Z. Für Pädagogische Psychol.* **2014**, *28*, 237–248. [CrossRef]

11. Steffensky, M.; Gold, B.; Holdynski, M.; Möller, K. Professional Vision of Classroom Management and Learning Support in Science Classrooms—Does Professional Vision Differ Across General and Content-Specific Classroom Interactions? *Int. J. Sci. Math. Educ.* **2015**, *13*, 351–368. [CrossRef]

12. Tolsdorf, Y.; Markic, S. Exploring Chemistry Student Teachers' Diagnostic Competence—A Qualitative Cross-Level Study. *Educ. Sci.* **2017**, *7*, 86. [CrossRef]

13. 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.

14. Schrader, F.-W. Diagnostische Kompetenz von Lehrpersonen [Teacher Diagnosis and Diagnostic Competence]. *Beiträge Zur Lehr.* **2013**, *31*, 154–165.

15. Artelt, C.; Rausch, T. Accuracy of teacher judgments: When and for What Reasons? In *Teacher's Professional Development: Assessment, Training, and Learning. The Future of Education Research*; Krolak-Schwerdt, S., Glock, S., Böhmer, M., Eds.; Sense Publishers: Rotterdam, The Netherlands, 2014; pp. 27–43.

16. Südkamp, A.; Kaiser, J.; Möller, J. Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *J. Educ. Psychol.* **2012**, *104*, 743–762. [CrossRef]

17. 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**, *57*, 175. [CrossRef]

18. Bromme, R. Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers [Competencies, functions and instructional behavior of the teacher]. In *Psychologie des Unterrichts und der Schule [Psychology of Teaching and School]*; Weinert, F.E., Ed.; Hogrefe: Göttingen, Germany, 1997; pp. 177–212.

19. Kaiser, G.; Blömeke, S.; König, J.; Busse, A.; Döhrmann, M.; Hoth, J. Professional competencies of (prospective) mathematics teachers—Cognitive versus situated approaches. *Educ. Stud. Math.* **2017**, *94*, 161–184. [CrossRef]

20. 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]

21. Hoth, J.; Kaiser, G.; Döhrmann, M.; König, J.; Blömeke, S. A Situated Approach to Assess Teachers' Professional Competencies Using Classroom Videos. In *Mathematics Teachers Engaging with Representations of Practice: A Dynamically Evolving Field*; Buchbinder, O., Kuntze, S., Eds.; ICME-13 Monographs; Springer: Berlin/Heidelberg, Germany, 2018; pp. 23–45.

22. 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]

23. Shulman, L.S. Those who understand: Knowledge growth in teaching. *Educ. Res.* **1986**, *15*, 4–14. [CrossRef]

24. Shulman, L.S. Knowledge and teaching: Foundations of the new reform. *Havard Educ. Rev.* **1987**, *57*, 1–22. [CrossRef]

25. 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]

26. 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]

27. Ball, D.L.; Thames, M.H.; Phelps, G. Content Knowledge for Teaching: What Makes It Special? *J. Teach. Educ.* **2008**, *59*, 389–407. [CrossRef]

28. 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.

29. 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]

30. Fischer, H.E.; Borowski, A.; Tepner, O. Professional Knowledge of Science Teachers. In *Second International Handbook of Science Education*; Fraser, B.J., Tobin, K., McRobbie, C.J., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 435–448, ISBN 978-1-4020-9040-0.

31. Baumert, J.; Kunter, M. The Effect of Content Knowledge and Pedagogical Content Knowledge on Instructional Quality and Student Achievement. 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. 175–205, ISBN 978-1-4614-5148-8.

32. Hill, H.C.; Schilling, S.G.; Ball, D.L. Developing Measures of Teachers' Mathematics Knowledge for Teaching. *Elem. Sch. J.* **2004**, *105*, 11–30. [CrossRef]

33. 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]

34. 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]

35. Kulgemeyer, C.; Riese, J. From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *J. Res. Sci. Teach.* **2018**, *55*, 1393–1418. [CrossRef]

36. 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 International Publishing: Cham, Switzerland, 2017; pp. 111–129, ISBN 978-3-319-50028-7.

37. Alonzo, A.C.; Kim, J. Affordances of video-based professional development for supporting physics teachers' judgments about evidence of student thinking. *Teach. Teach. Educ.* **2018**, *76*, 283–297. [CrossRef]

38. 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]

39. Sorge, S.; Stender, A.; Neumann, K. The Development of Science Teachers' Professional Competence. In *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*; Hume, A., Cooper, R., Borowski, A., Eds.; Springer: Singapore, 2019; pp. 149–164.

40. Codreanu, E.; Sommerhoff, D.; Huber, S.; Ufer, S.; Seidel, T. Between authenticity and cognitive demand: Finding a balance in designing a video-based simulation in the context of mathematics teacher education. *Teach. Teach. Educ.* **2020**, *95*. [CrossRef]

41. Seidel, T.; Stürmer, K. Modeling and measuring the structure of professional vision in preservice teachers. *Am. Educ. Res. J.* **2014**, *51*, 739–771. [CrossRef]

42. Sherin, M.G.; van Es, E.A. Effects of Video Club Participation on Teachers' Professional Vision. *J. Teach. Educ.* **2009**, *60*, 20–37. [CrossRef]

43. van Es, E.A.; Sherin, M.G. Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *J. Technol. Teach.* **2002**, *10*, 571–596.

44. Ingenkamp, K.; Lissmann, U. *Lehrbuch der Pädagogischen Diagnostik [Textbook of Pedagogical Diagnostics]*, 6th ed.; Beltz: Weinheim, Germany, 2008; ISBN 3407255039.

45. Fischer, F.; Kollar, I.; Ufer, S.; Sodian, B.; Hussmann, H.; Pekrun, R.; Neuhaus, B.J.; Dorner, B.; Pankofer, S.; Fischer, M.R.; et al. Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learn. Res.* **2014**, *5*, 28–45. [CrossRef]

46. 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: Abingdon, UK, 2016; pp. 77–94, ISBN 9781315749136.

47. Barendsen, E.; Henze, I. Relating Teacher PCK and Teacher Practice Using Classroom Observation. *Res. Sci. Educ.* **2019**, *49*, 1141–1175. [CrossRef]

48. Lehane, L.; Bertram, A. Getting to the CoRe of it: A review of a specific PCK conceptual lens in science educational research. *Educ. Química* **2016**, *27*, 52–58. [CrossRef]

49. Sherin, M.G.; Russ, R.S.; Sherin, B.L.; Colestock, A. Professional Vision in Action: An Exploratory Study. *Issues Teach. Educ.* **2008**, *17*, 27–46.

50. McElvany, N.; Schroeder, S.; Hachfeld, A.; Baumert, J.; Richter, T.; Schnotz, W.; Horz, H.; Ullrich, M. Diagnostische Fähigkeiten von Lehrkräften [Diagnostic skills of teachers]. *Z. Für Pädagogische Psychol.* **2009**, *23*, 223–235. [CrossRef]

51. Helmke, A.; Schrader, F.-W. Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teach. Teach. Educ.* **1987**, *3*, 91–98. [CrossRef]

52. Carter, K.; Cushing, K.; Sabers, D.; Stein, P.; Berliner, D.C. Expert-Novice Differences in Perceiving and Processing Visual Classroom Information. *J. Teach. Educ.* **1988**, *39*, 25–31. [CrossRef]

53. 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]

54. 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 Educ.* **2020**, *7*. [CrossRef]

55. Todorova, M.; Sunder, C.; Steffensky, M.; Möller, K. Pre-service teachers' professional vision of instructional support in primary science classes: How content-specific is this skill and which learning opportunities in initial teacher education are relevant for its acquisition? *Teach. Teach. Educ.* **2017**, *68*, 275–288. [CrossRef]

56. 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]

57. Kersting, N.B.; Givvin, K.B.; Sotelo, F.L.; Stigler, J.W. Teachers' Analyses of Classroom Video Predict Student Learning of Mathematics: Further Explorations of a Novel Measure of Teacher Knowledge. *J. Teach. Educ.* **2010**, *61*, 172–181. [CrossRef]

58. Kaiser, J.; Helm, F.; Retelsdorf, J.; Südkamp, A.; Möller, J. Zum Zusammenhang von Intelligenz und Urteilsgenauigkeit bei der Beurteilung von Schülerleistungen im Simulierten Klassenraum [On the Relation of Intelligence and Judgment Accuracy in the Process of Assessing Student Achievement in the Simulated Classroom]. *Z. Für Pädagogische Psychol.* **2012**, *26*, 251–261. [CrossRef]

59. Machts, N.; Kaiser, J.; Schmidt, F.T.C.; Möller, J. Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educ. Res. Rev.* **2016**, *19*, 85–103. [CrossRef]

60. de Bruin, A.B.; Schmidt, H.G.; Rikers, R.M.J.P. The role of basic science knowledge and clinical knowledge in diagnostic reasoning: A structural equation modeling approach. *Acad. Med.* **2005**, *80*, 765–773. [CrossRef]

61. Braun, L.T.; Zottmann, J.M.; Adolf, C.; Lottspeich, C.; Then, C.; Wirth, S.; Fischer, M.R.; Schmidmaier, R. Representation scaffolds improve diagnostic efficiency in medical students. *Med. Educ.* **2017**, *51*, 1118–1126. [CrossRef]

62. Berliner, D.C. Learning about and learning from expert teachers. *Int. J. Educ. Res.* **2001**, *35*, 463–482. [CrossRef]

63. Beck, E.; Baer, M.; Guldimann, T.; Bischoff, S.; Brühwiler, C.; Müller, P.; Niedermann, R.; Rogalla, M.; Vogt, F. *Adaptive Lehrkompetenz. Analyse und Struktur, Veränderbarkeit und Wirkung Handlungssteuernden Lehrerwissens [Adaptive Teaching Competence. Analysis and Structure, Changeability and Effect of Action-Guiding Teacher Knowledge]*; Waxmann: Münster, Germany, 2008.

64. Brühwiler, C. *Adaptive Lehrkompetenz und Schulisches Lernen. Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler [Adaptive Teaching Competence and School-Based Learning. Effects of Teachers' Action-Guiding Cognitions on Instructional Processes and Student Learning Outcomes]*; Waxmann: Münster, Germany, 2014.

65. Behling, F.; Förtsch, C.; Neuhaus, B.J. Sprachsensibler Biologieunterricht–Förderung professioneller Handlungskompetenz und professioneller Wahrnehmung durch videogestützte live-Unterrichtsbeobachtung. Eine Projektbeschreibung [Language-sensitive Biology Instruction–Fostering Professional Competence and Professional Vision Through Video-based Live Lesson Observation. A Project Description]. *Z. Didakt. Naturwissenschaften* **2019**, *42*, 271. [CrossRef]

66. Goldman, E.; Barron, L. Using Hypermedia to Improve the Preparation of Elementary Teachers. *J. Teach. Educ.* **1990**, *41*, 21–31. [CrossRef]

67. Grossman, P.; Compton, C.; Igra, D.; Ronfeldt, M.; Shahan, E.; Williamson, P.W. Teaching Practice: A Cross-Professional Perspective. *Teach. Coll. Rec.* **2009**, *111*, 2055–2100.

68. Wedel, A.; Müller, C.R.; Pfetsch, J.; Ittel, A. Training teachers' diagnostic competence with problem-based learning: A pilot and replication study. *Teach. Teach. Educ.* **2019**, *86*, 1–14. [CrossRef]

69. Wiens, P.D.; Beck, J.S.; Lunsmann, C.J. Assessing teacher pedagogical knowledge: The Video Assessment of Teacher Knowledge (VATK). *Educ. Stud.* **2020**, 1–17. [CrossRef]

70. Ohle, A.; McElvany, N. Teacher diagnostic competences and their practical relevance. Special issue editorial. *J. Educ. Res. Online* **2015**, *7*, 5–10.

71. Spector, P.E. Do Not Cross Me: Optimizing the Use of Cross-Sectional Designs. *J. Bus. Psychol.* **2019**, *34*, 125–137. [CrossRef]

72. Südkamp, A.; Praetorius, A.-K. (Eds.) *Diagnostische Kompetenz von Lehrkräften*; Waxmann: Münster, Germany, 2017.

73. 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.

74. Campbell, D.E. How to write good multiple-choice questions. *J. Paediatr. Child Health* **2011**, *47*, 322–325. [CrossRef]

75. Jüttner, M.; Neuhaus, B.J. Validation of a Paper-and-Pencil Test Instrument Measuring Biology Teachers' Pedagogical Content Knowledge by Using Think-Aloud Interviews. *J. Educ. Train. Stud.* **2013**, *1*, 113–125. [CrossRef]

76. 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. Assess. Eval. Acc.* **2013**, *25*, 45–67. [CrossRef]

77. 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]

78. 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*, 7–28.

79. 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, Alemanha, 2002; ISBN 3801716465.

80. Kunina-Habenicht, O.; Maurer, C.; Wolf, K.; Kunter, M.; Holzberger, D.; Schmidt, M.; Seidel, T.; Dicke, T.; Teuber, Z.; Koc-Januchta, M.; 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]

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 23 February 2021).

82. 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.

83. 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]

84. Bond, T.G.; Fox, C.M.F. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd ed.; Lawrence Erlbaum: Mahwah, NJ, USA, 2007.

85. Boone, W.J.; Staver, J.; Yale, M. *Rasch Analysis in the Human Sciences*; Springer: Dordrecht, The Netherlands, 2014; ISBN 9789400768574.

86. Linacre, J.M. A User's Guide to Winsteps®Ministeps Rasch-Model Computer Programs. Program Manual 4.8.0. Available online: https://www.winsteps.com/a/Winsteps-Manual.pdf (accessed on 23 February 2021).

87. Boone, W.J.; Staver, J. *Advances in Rasch Analysis in the Human Science*; Springer: Dordrecht, The Netherlands, 2020.

88. Wright, B.D.; Linacre, J.M. Reasonable mean-square fit values. *Rasch Meas. Trans.* **1994**, *8*, 370.

89. Questback GmbH. *EFS Survey*; Questback GmbH: Köln, Germany, 2018.

90. Dorfner, T.; Förtsch, C.; Neuhaus, B.J. Die methodische und inhaltliche Ausrichtung quantitativer Videostudien zur Unterrichtsqualität im mathematisch-naturwissenschaftlichen Unterricht: Ein Review [The methodical and content-related orientation of quantitative video studies on instructional quality in mathematics and science education]. *ZfDN* **2017**, *23*, 261–285. [CrossRef]

91. Kramer, M.; Förtsch, C.; Seidel, T.; Neuhaus, B.J. Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies Educ. Eval.* **2021**, *28*. [CrossRef]

92. Scheiner, T. Teacher noticing: Enlightening or blinding? *ZDM Math. Educ.* **2016**, *48*, 227–238. [CrossRef]

93. Vogt, F.; Schmiemann, P. Assessing Biology Pre-Service Teachers' Professional Vision of Teaching Scientific Inquiry. *Educ. Sci.* **2020**, *10*, 332. [CrossRef]

94. Kaiser, G.; Blömeke, S.; Busse, A.; Döhrmann, M.; König, J. Professional knowledge of (prospective) Mathematics teachers—Its structure and development1. 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.

95. Arbuckle, J.L. *Amos (Version 26.0)*; SPSS: Chicago, IL, USA, 2019.

96. Hu, L.-T.; Bentler, P.M. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol. Methods* **1998**, *3*, 424–453. [CrossRef]

97. Finger, R.P.; Fenwick, E.; Pesudovs, K.; Marella, M.; Lamoureux, E.L.; Holz, F.G. Rasch analysis reveals problems with multiplicative scoring in the macular disease quality of life questionnaire. *Ophthalmology* **2012**, *119*, 2351–2357. [CrossRef] [PubMed]