Investigating pre-service biology teachers' diagnostic competences with a video-based assessment tool

Effects of instructional support on professional knowledge as one component of diagnostic competences

Dissertation der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

Maria Kramer

München, 2021

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. Birgit J. Neuhaus
im Bereich Didaktik der Biologie
an der Ludwig-Maximilians-Universität München.

Die vorliegende Arbeit wurde durch die Deutsche Forschungsgemeinschaft (DFG, Förderkennzeichen: NE 1196/8-1, FOR 2385) gefördert. Die Arbeit wurde am Lehrstuhl Didaktik der Biologie der Ludwig-Maximilians-Universität durchgeführt.

1. Gutachterin:	Prof. Dr. Birgit. J. Neuhaus
2. Gutachterin:	Prof. Dr. Tina Seidel
Datum der Abgabe:	25.03.2021
Datum der mündlichen Prüfung:	15.07.2021

Statutory declaration and statement

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 25.03.2021 *Maria Kramer*

Erklärung

Hiermit erkläre ich,

dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

München, den 25.03.2021 *Maria Kramer*

Abstract

In order to be able to recognize and cope with critical situations of teaching and learning, teachers need diagnostic competences, which should already be fostered in teacher education at university. Diagnostic competences can be defined as teachers' dispositions enabling them to apply their knowledge in diagnostic activities according to professional standards to collect and interpret data accurately in order to make appropriate decisions. In this definition, three components of diagnostic competences become apparent: the professional knowledge base, diagnostic activities, and diagnostic accuracy. The professional knowledge base consists of the facets pedagogical-psychological knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK). Pre-service teachers have to be able to apply knowledge in diagnostic contexts that are real or close to everyday life by using diagnostic activities. Diagnostic activities describe those activities that teachers execute to generate and evaluate data in order to produce a diagnosis. Additionally, their diagnosis has to meet qualitative demands in terms of diagnostic accuracy, which describes the adequacy of a person's diagnosis in terms of objective criteria. It has not yet been systematically investigated how the components of diagnostic competences are interrelated. Furthermore, there is a growing debate about how to foster knowledge development in higher education. The investigation of different instructional approaches of knowledge presentation (e.g., by a lecturer, via texts, via video-based tools) can be connected to this debate. With regard to the demand to investigate professional knowledge as well as diagnostic activities using real-life diagnostic contexts, video-based tools are considered promising. Video-based tools can provide an authentic representation of the subject-specific professional everyday life of teachers and their tasks. That makes it possible to measure and promote specific components of diagnostic competences in a situated way. For biology teacher education, there is a lack of tools in which subject-specific relevant situations from the biology classroom are the subject of diagnosis.

To close this gap, the video-based assessment tool DiKoBi *Assess* was developed in the project COSIMA (*Facilitating diagnostic competences in simulation-based learning environments in higher education*), which focused on the facilitation of professional knowledge and diagnostic competences of biology teachers and was funded by the DFG (German Research Foundation, NE-1196/8-1). In DiKoBi *Assess* (German acronym for *diagnostic competences of biology teachers in biology classrooms*), biology-specific challenges of different classroom situations have to be diagnosed with regard to dimensions

of subject-specific instructional quality (i.e., *level of students' cognitive activities and creation of situational interest*, *dealing with (specific) student ideas and errors*, *use of technical language*, *use of experiments*, *use of models*, and *conceptual instruction*).

Overall, the present dissertation had three aims: (1) to validate the video-based tool DiKoBi *Assess* with respect to the integrated biology-specific challenges, the tasks used, and the construct of *diagnostic activities* (i.e., the activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*) used for analysis. The validation process is completed by comparing the construct of *diagnostic activities* with the construct of *professional vision* (regarding the included aspects *description*, *explanation*, and *prediction*) established in video-based teacher research. Further aims were (2) the investigation of the relationships between the three components that constitute diagnostic competences, and (3) the analysis of the effects of different types of instructional support on professional knowledge as one component of diagnostic competences.

The three aims were addressed in the first funding phase of the COSIMA project. Following a preliminary study with five in-service biology teachers, in which the validity of the video-based tool was tested, two main studies with pre-service biology teachers took place. Data was collected in both Study 1 ($N = 90$ pre-service teachers) and Study 2 ($N = 103$ pre-service teachers) in a pre-post design. In both studies, professional knowledge was assessed with paper-pencil tests, diagnostic activities and diagnostic accuracy were assessed with the video-based tool DiKoBi *Assess*.

The validation of DiKoBi *Assess* showed that the embedded biology-specific classroom situations were perceived as authentic, that biology-specific challenges in the respective classroom situations were diagnosable by in-service biology teachers, and that the integrated tasks triggered specifically the three targeted diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*. Thus, with DiKoBi *Assess*, specific diagnostic activities can be triggered and measured. Moreover, as a result of comparing the constructs *diagnostic activities* and *professional vision*, it was possible to identify and describe categories relevant for diagnosing the biology-specific classroom situations presented in DiKoBi *Assess*. The identified categories could then be used for analysis in the subsequent main studies. The evaluation of the two main studies showed that especially PCK was important for the application of the diagnostic activities and for diagnostic accuracy. In addition, a relationship between PK and diagnostic activities was found. Both main studies also indicated an impact of the work with DiKoBi *Assess* on the development of PCK. Furthermore, the use of the video-based tool in combination with instructional support provided by a lecturer can be considered effective. While for the development of CK, a non-integrated instruction might be beneficial, PCK benefited slightly more from the integrated instruction of the knowledge facets. However, the results suggest that diagnosing instructional quality is more effective for knowledge development than instructional approaches such as text interventions, but that in terms of PCK development, for example, greater linking of knowledge facets may also be advisable when instructing curricular content.

Overall, it can be concluded that DiKoBi *Assess* is a valid measurement instrument for studies on diagnostic competences of pre-service biology teachers. Furthermore, the results of this dissertation suggest that working with DiKoBi *Assess* and diagnosing instructional quality can promote subject-specific knowledge (PCK), which in turn is a critical component of biology teachers' diagnostic competences and therefore requires explicit support. Existing instructional methods of teacher education, in which knowledge acquisition is promoted via texts or lectures, should therefore be enriched with practice-oriented examples. This can be done via video-based tools that could be conducive to the development and stronger interlocking of the more practice-oriented knowledge facets PCK and PK. In addition, such tools can also be beneficial to address the knowledge facets in an interrelated way and to enhance linking between the knowledge facets. However, explicit specialized seminars addressing subject content still seem useful for developing a sound CK base.

Zusammenfassung

Um entscheidende Lehr- und Lernsituationen erkennen und bewältigen zu können, benötigen Lehrkräfte Diagnosekompetenzen, die bereits im Rahmen der Lehrerausbildung an der Hochschule gefördert werden sollten. Diagnosekompetenzen können als Dispositionen von Lehrkräften definiert werden, die es ihnen ermöglichen, ihr Wissen in diagnostischen Aktivitäten gemäß professionellen Standards anzuwenden, um Daten zu sammeln und korrekt zu interpretieren und somit angemessene Entscheidungen zu treffen. In dieser Definition werden drei Komponenten von Diagnosekompetenzen sichtbar: eine professionelle Wissensbasis, diagnostische Aktivitäten und diagnostische Akkuratheit. Die professionelle Wissensbasis besteht aus den Facetten pädagogisch-psychologisches Wissen (PK), Fachwissen (CK) und fachdidaktisches Wissen (PCK). Die angehenden Lehrkräfte müssen in der Lage sein, ihr Wissen innerhalb realer bzw. alltagsnaher Diagnosekontexte in diagnostischen Aktivitäten anzuwenden. Diagnostische Aktivitäten beschreiben jene Aktivitäten, die Lehrer ausführen, um Daten zu generieren und zu evaluieren, damit eine Diagnose erstellt werden kann. Zusätzlich muss die Diagnose qualitativen Anforderungen genügen, beispielsweise hinsichtlich ihrer Akkuratheit, womit die Adäquatheit der Diagnose einer Person hinsichtlich objektiver Kriterien beschrieben wird. Wie die Komponenten der Diagnosekompetenzen zusammenhängen, wurde bisher noch nicht systematisch untersucht. Darüber hinaus kann mit Blick auf die universitäre Ausbildung eine stärker werdende Debatte hinsichtlich der Förderung der professionellen Wissensbasis ausgemacht werden, an die sich die Untersuchung verschiedener instruktionaler Ansätze der Wissensvermittlung (z. B. durch Vorlesungen, Textarbeit, über videobasierte Tools) anschließen kann. Im Hinblick auf die Forderung, dieses Wissen sowie entsprechende diagnostische Aktivitäten anhand realer Diagnosekontexte zu untersuchen, gelten videobasierte Tools als vielversprechend. Der Einsatz videobasierter Tools wird als eine Möglichkeit angesehen, um den fachspezifischen Berufsalltag möglichst authentisch abzubilden und somit spezifische Komponenten von Diagnosekompetenzen situiert messen und fördern zu können. Für die Ausbildung angehender Biologielehrkräfte fehlen bisher größtenteils Tools, in denen fachspezifisch relevante Situationen aus dem Biologieunterricht Gegenstand der Diagnose sind.

Um diese Lücke zu schließen, wurde in dem von der DFG (Deutsche Forschungsgemeinschaft) geförderten Projekt COSIMA (*Förderung von Diagnosekompetenzen in simulationsbasierten Lernumgebungen in der Hochschule*), in dem die Förderung von Professionswissen und diagnostischen Kompetenzen von Biologielehrkräften im Mittelpunkt

steht (NE-1196/8-1), das videobasierte Diagnose-Tool DiKoBi *Assess* entwickelt. In DiKoBi *Assess* (Akronym für *Diagnosekompetenzen von Biologielehrkräften im Biologieunterricht*) sollen biologiespezifische Herausforderungen verschiedener Unterrichtssituationen im Hinblick auf die fachspezifische Unterrichtsqualität diagnostiziert werden (mit Bezug auf das Niveau der kognitiven Aktivitäten der Lernenden und der Erzeugung von situativem Interesse, den Umgang mit (spezifischen) Schülerideen und Fehlern, der Verwendung von Fachsprache, den Einsatz von Experimenten, die Verwendung von Modellen, und konzeptorientiertem Unterricht).

Die vorliegende Dissertation verfolgt insgesamt drei Ziele: (1) die Validierung des videobasierten Tools DiKoBi *Assess* hinsichtlich der integrierten biologiespezifischen Herausforderungen, der verwendeten Aufgaben sowie des für die Analyse verwendeten Konstrukts der *diagnostischen Aktivitäten* (insbesondere die Aktivitäten *Evidenzen generieren*, *Evidenzen evaluieren* und *Schlussfolgerungen ziehen*). Abgeschlossen wird der Validierungsprozess durch einen Vergleich des Konstrukts der *diagnostischen Aktivitäten* mit dem in der videobasierten Lehrerforschung etablierten Konstrukt *Professional Vision* (hinsichtlich der enthaltenen Aspekte *Beschreibung*, *Erklärung* und *Vorhersage*). Weiterhin verfolgt werden (2) die Untersuchung der Zusammenhänge zwischen den drei Komponenten von Diagnosekompetenzen sowie (3) die Analyse der Effekte unterschiedlicher instruktionaler Unterstützungsmaßnahmen auf die professionelle Wissensbasis als Komponente von Diagnosekompetenzen.

Die drei Ziele wurden im Rahmen der ersten Förderphase des COSIMA Projektes bearbeitet. Anschließend an eine Vorstudie mit fünf praktizierenden Biologielehrkräften, in der die Validität des videobasierten Tools überprüft wurde, fanden zwei Hauptstudien mit angehenden Biologielehrkräften statt. Daten wurden sowohl in Studie 1 ($N = 90$ angehende Biologielehrkräfte) als auch in Studie 2 ($N = 103$ angehende Biologielehrkräfte) in einem Prä-Post-Design erhoben. In beiden Studien wurde das Professionswissen mit Paper-Pencil-Tests, die diagnostischen Aktivitäten und die diagnostische Akkuratheit mit dem videobasierten Tool DiKoBi *Assess* erfasst.

Die Validierung von DiKoBi *Assess* zeigte, dass die eingebetteten fachspezifischen Unterrichtssituationen als authentisch wahrgenommen wurden, dass biologiespezifische Herausforderungen in den jeweiligen Unterrichtssituationen durch praktizierende Biologielehrkräfte diagnostizierbar waren, und dass die integrierten Aufgaben genau die drei anvisierten diagnostischen Aktivitäten *Evidenzen generieren*, *Evidenzen evaluieren* und *Schlussfolgerungen ziehen* triggerten. Somit können mit DiKoBi *Assess* gezielt diagnostische

Aktivitäten ausgelöst und gemessen werden. Zudem konnten als Resultat aus dem Vergleich der Konstrukte *diagnostische Aktivitäten* und *Professional Vision* Kategorien identifiziert und beschrieben werden, die für das Diagnostizieren der biologiespezifischen Unterrichtssituationen aus DiKoBi *Assess* relevant waren und somit in den nachfolgenden Hauptstudien als Analysekategorien genutzt werden konnten. Die Auswertung der Hauptstudien zeigte, dass vor allem PCK wichtig für die Anwendung der diagnostischen Aktivitäten und die diagnostische Akkuratheit war. Daneben bestand auch ein Zusammenhang zwischen PK und den diagnostischen Aktivitäten. Beide Hauptstudien deuteten zudem auf einen Einfluss der Arbeit in DiKoBi *Assess* für den Aufbau von PCK hin. Als sinnvoll kann darüber hinaus der Einsatz des videobasierten Tools in Kombination mit dozentenbasierter Instruktion angesehen werden. Während für die Entwicklung von CK eine nicht-integrierte Vermittlung der Wissensfacetten förderlich zu sein scheint, profitiert PCK deskriptiv etwas stärker von einer integrierten Vermittlung. Insgesamt legen die Ergebnisse nahe, dass das Diagnostizieren biologiespezifischer Unterrichtssituationen in DiKoBi *Assess* effektiver für den Wissensaufbau als instruktionale Ansätze wie z.B. Textinterventionen ist, dass aber hinsichtlich der Förderung von PCK auch eine stärkere Verknüpfung der Wissensfacetten bei der Instruktion von curricularen Inhalten angebracht ist.

Daher kann hinsichtlich der Ergebnisse dieser Dissertation festgehalten werden, dass DiKoBi *Assess* ein valides Messinstrument für Untersuchungen zu Diagnosekompetenzen angehender Biologielehrkräfte ist. Darüber hinaus kann durch die Arbeit mit DiKoBi *Assess* und dem Diagnostizieren von Unterrichtsqualität fachspezifisches Wissen (PCK) gefördert werden, welches wiederum eine entscheidende Komponente der diagnostischen Kompetenzen von Biologielehrkräften darstellt und daher expliziter Förderung bedarf. Bisherige Lehrmethoden der Lehrerausbildung, in denen Wissen über Texte oder Vorlesungen vermittelt wird, sollten daher vor allem im Hinblick auf die Förderung der anwendungsorientierten Wissensfacetten PCK und PK noch stärker um Anwendungen mit Praxisbezug ergänzt werden. Dies kann beispielweise über videobasierte Tools erfolgen. Zusätzlich können derartige Tools auch hilfreich sein, um die Wissensfacetten in integrierter Weise zu adressieren und eine stärkere Vernetzung zwischen den Wissensfacetten zu ermöglichen. Explizite Fachseminare, in denen spezifische Fachinhalte vermittelt werden, erscheinen jedoch nach wie vor sinnvoll, um eine fundierte CK-Basis zu entwickeln.

Content

Abbreviations

BilWiss	Educational Knowledge and the Development of Professional Competence in Teacher Training
CK	content knowledge
CLASS-S	Classroom Assessment Scoring System-Secondary
COACTIV	Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers
COSIMA	Facilitating Diagnostic Competences in Simulation-based Learning Environments in Higher Education
DiaCoM	Explaining Teachers' Diagnostic Judgements by Cognitive Modeling
DFG	German Research Foundation
DiKoBi	Diagnostic Competences of Biology Teachers in Biology Classrooms
KiL	Measuring the professional knowledge of preservice mathematics and science teachers
KMK	Standing Conference of the Ministers of Education and Cultural Affairs of the Länder
MT21	Mathematics Teaching in the 21st Century
QuIP	Quality of Instruction in Physics
PCK	pedagogical content knowledge
cPCK	collective PCK
ePCK	enacted PCK
pPCK	personal PCK
PID	Perceiving, Interpreting, Decision-making
PK	pedagogical knowledge
PST	pre-service teacher
ProwiN	Professional Knowledge of Teachers in Science
RCM	Refined Consensus Model
SCR^{Bio}	Simulated Classroom Biology
TEDS-M	Teacher Education and Development Study

List of publications

Publications of the cumulative dissertation

Publication I	**Kramer, M.**, Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, *7*(1). https://doi.org/10.1080/2331186X.2020.1823155
Publication II	**Kramer, M.**, Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation*, *28*. https://doi.org/10.1016/j.stueduc.2020.100973
Publication III	**Kramer, M.**, Förtsch, C., Boone, W. J., Seidel, T., & Neuhaus, B. J. (2021). Investigating Pre-Service Biology Teachers' Diagnostic Competences: Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy. *Education Sciences*, *11*(3), 89. https://doi.org/10.3390/educsci11030089
Manuscript I	**Kramer, M**., Förtsch, C., & Neuhaus, B. J. (moderate revision). Integrating or not-integrating – that is the question. Effects of integrated instruction on the development of pre-service biology teachers' professional knowledge. *Frontiers in Education*.
Manuscript II	**Kramer, M**., Förtsch, C., & Neuhaus, B. J. (under revision). Can pre-service biology teachers' professional knowledge and diagnostic activities be fostered by self-directed knowledge acquisition via texts? *Education Sciences*.

Declaration of contribution as a co-author

Erklärung über die Eigenanteile bei Ko-Autorenschaft

Hiermit wird bestätigt, dass die folgenden drei Publikationen und die zwei Manuskripte federführend von Frau Maria Kramer im Rahmen ihrer Dissertation abgefasst wurden. Dies geschah mit folgenden Anteilen:

Publication I

Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, *7*(1).
https://doi.org/10.1080/2331186X.2020.1823155

Frau Maria Kramer hat ein bereits im Projekt vorhandenes Kodiermanual erheblich ausgebaut und weiter entwickelt, die Implementierung des videobasierten Diagnose-Tools DiKoBi *Assess* auf der Online-Plattform Unipark (QuestBack GmbH) fertiggestellt, bereits erhobene Daten für die Validierung von DiKoBi *Assess* ausgewertet, Validierungsergebnisse aufbereitet, DiKoBi *Assess* entsprechend der Validierungsergebnisse und im Hinblick auf weitere Studien angepasst, den Validierungsartikel konzipiert und ihn federführend geschrieben.

Die Koautorin Julia Stürmer entwickelte die Erstfassung des Kodiermanuals, konzipierte und videografierte verschiedene Videosituationen, entwickelte aus diesen das Tool DiKoBi *Assess*, implementierte die Grundzüge des Tools auf der Online-Plattform Unipark (QuestBack GmbH) und erhob die Daten für die Validierung. Die Koautorin Sonja Förtsch videografierte ebenfalls einen Teil der Videosituationen und wirkte sowohl bei der Datenerhebung als auch –auswertung mit. Der Koautor Christian Förtsch und die Koautorin Birgit J. Neuhaus entwickelten das Studiendesign, wirkten bei der Weiterentwicklung der Kodiermanuale mit, unterstützten bei der Datenerhebung und -auswertung und trugen substanziell zur Konzipierung und Überarbeitung der Publikation bei. Die Koautorin Tina Seidel unterstütze bei der Datenauswertung sowie bei der Konzipierung und Überarbeitung der Publikation.

Publication II

Kramer, M., Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing Two Constructs for Describing and Analyzing Teachers' Diagnostic Processes. *Studies in Educational Evaluation*, *28*. https://doi.org/10.1016/j.stueduc.2020.100973

Frau Maria Kramer hat ein weiteres Kodiermanual entwickelt, bereits vorhandene Daten kodiert und ausgewertet, die Daten für die Publikation aufgearbeitet, den Artikel konzipiert und ihn federführend geschrieben.

Die beiden Koautorinnen und der Koautor unterstützten bei der Datendeutung und trugen substanziell zur Überarbeitung der Publikation bei.

Publication III

Kramer, M., Förtsch, C., Boone, W. J., Seidel, T., & Neuhaus, B. J. (2021). Investigating Pre-Service Biology Teachers' Diagnostic Competences: Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy. *Education* Sciences, *11*(3), 89. https://doi.org/10.3390/educsci11030089

Frau Maria Kramer hat die Studie federführend durchgeführt, die Kodiermanuale weiterentwickelt, die Daten statistisch ausgewertet, die Daten für die Publikation aufgearbeitet, den Artikel konzipiert und ihn federführend geschrieben.

Die beiden Koautorinnen und der Koautor Christian Förtsch entwickelten das Studiendesign mit, unterstützten bei der Datenerhebung und -deutung und trugen substanziell zur Überarbeitung der Publikation bei. Der Koautor William J. Boone unterstützte bei der statistischen Datenauswertung und -aufarbeitung und wirkte an der Überarbeitung der Publikation mit.

Manuscript I

Kramer, M., Förtsch, C., & Neuhaus, B. J. (moderate revision). Integrating or not-integrating – that is the question. Effects of integrated instruction on the development of pre-service biology teachers' professional knowledge. *Frontiers in Education*.

Frau Maria Kramer hat das Material für die Intervention konzipiert, die Studie federführend durchgeführt, die Kodiermanuale weiterentwickelt, die Daten statistisch ausgewertet, die Daten für die Publikation aufgearbeitet, den Artikel konzipiert und ihn federführend geschrieben.

Die Koautorin und der Koautor entwickelten das Studiendesign, unterstützten bei der Materialentwicklung, Datenerhebung und -deutung und trugen substanziell zur Überarbeitung der Publikation bei.

Manuscript II

Kramer, M., Förtsch, C., & Neuhaus, B. J. (under revision). Can pre-service biology teachers' professional knowledge and diagnostic activities be fostered by self-directed knowledge acquisition via texts? *Education Sciences*.

Frau Maria Kramer hat das Material für die Intervention konzipiert, die Studie federführend durchgeführt, die Kodiermanuale weiterentwickelt, die Daten statistisch ausgewertet, die Daten für die Publikation aufgearbeitet, den Artikel konzipiert und ihn federführend geschrieben.

Die Koautorin und der Koautor entwickelten das Studiendesign, unterstützten bei der Materialentwicklung, Datenerhebung und -deutung und trugen substanziell zur Überarbeitung der Publikation bei.

München, den 25.03.2021 *Maria Kramer*

München, den 25.03.2021 *Prof. Dr. Birgit J. Neuhaus*

1. Introduction

As part of their professional competence, teachers must solve a wide variety of problem situations. Therefore, they engage in reasoning processes to make decisions that should lead to effective classroom instruction and instructional quality that, in turn, improves student achievement (Baumert & Kunter, 2013b; Seidel & Shavelson, 2007). Similar to a physician's behavior (identifying and interpreting symptoms of a patient to decide how best to treat the patient), a teacher's assessment of classroom situations and incidents can be understood as a diagnosis aimed at adapting instruction to best meet students' needs (Heitzmann et al., 2019; Klug et al., 2013). In order to effectively manage such diagnostic contexts, teachers need diagnostic competences[1] that can be "defined 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 high-quality decisions" (Heitzmann et al., 2019, p. 5). This definition originates from a joint research project of actors in medical and teacher education, who are investigating diagnostic competences in their respective professions in the DFG-funded project COSIMA (*Facilitating diagnostic competences in simulation-based learning environments in higher education*, FOR 2385). The transfer of previous findings and constructs of diagnosis from medicine to constructs of teacher education represents a central research approach of the COSIMA group. According to the definition of diagnostic competences, three components have been described that need to be considered when investigating teachers' diagnostic competences: teachers' professional knowledge as part of their cognitive dispositions, diagnostic activities that teachers engage in to derive a diagnosis, and a quality measure that can be realized in terms of accuracy to achieve adequate follow-up decisions (Heitzmann et al., 2019; Chernikova et al., in press). Since applying knowledge to individual cases is seen as challenging not only for pre-service teachers but also for young professionals (Berliner, 2001; Heitzmann et al., 2019), there is a demand to promote pre-service teachers' knowledge development, and thus, their diagnostic competences already in teacher education at university by providing opportunities for diagnosing. This approach is emphasized by researchers because competences are acquired through experience and learning in relevant, challenging situations and can be influenced by external interventions

[1] Throughout this dissertation, the plural "diagnostic competences" is used to indicate that there is no global construct of diagnostic competence but that its conceptualization depends on the specific focus of the study, the diagnostic occasion, or the level of the diagnosis (cf. Karst et al., 2014, McElvany et al. 2009, Steffensky et al., 2015).

and additional instructional support (Heitzmann et al., 2019; Janssen & Lazonder, 2016; Klieme et al., 2008; Weinert, 2001). Therefore, video-based programs and tools have been developed that approximate practice and allow for the application of knowledge while considering the specificity of a situation (Gaudin & Chaliès, 2015; Gold & Holodynski, 2017; Hoth et al., 2018). In science education, there are a few video-based tools that provide diagnostic situations, but whole-class scenarios in which science-specific features of instructional quality have to be assessed for diagnosis have not explicitly been designed yet. Instead, the diagnostic foci of the tools are directly related to student-teacher-interaction, student behavior and student learning (e.g., Codreanu et al., 2020; Hoth et al., 2016; Kaiser et al., 2015; Kersting, 2008; Santagata et al., 2007), but not to specific instructional quality dimensions such as cognitive activation that affect student achievement (Förtsch et al., 2016). Since it is the subject-specific dimensions in particular that are crucial for instructional quality and student achievement in the subject (Helmke & Schrader, 1987; Karing et al., 2011; Seidel & Shavelson, 2007), an instrument that allows the assessment of subject-specific instructional quality with consideration of the teachers' instructional behavior was developed and validated: the video-based tool DiKoBi *Assess*[2] (German acronym for *diagnostic competences of biology teachers in biology classrooms*). This tool is intended to investigate questions regarding the analysis and promotion of the three components of diagnostic competences within biology education.

To frame the publications and manuscripts as well as the theoretical approaches and constructs used therein, the introduction and discussion of this dissertation are structured as follows: In order to understand the structure and conception of the video-based tool DiKoBi *Assess* and the associated diagnostic context, the introduction starts with the description of generic and subject-specific dimensions of instructional quality that underlie effective teaching (see Section 1.1). This perspective representing research on teaching effectiveness is then complemented by aspects concerning research on teacher professionalism. Based on the situated approach of modeling the professional competence of teachers as a continuum (see Section 1.2), the components of diagnostic competences are derived from previous research traditions before they are brought together and embedded in a conceptual framework of the COSIMA research project, whose evaluation with regard to the conceptualization of diagnostic competences and possible effects of instructional factors are still pending (see

[2] The video-based tool DiKoBi can be used in two ways: (1) as a measuring instrument (DiKoBi *Assess*) and (2) as a learning environment (DiKoBi *Learn*). This dissertation is mainly concerned with the tool's use as a measurement instrument and therefore refers to the tool as DiKoBi *Assess*.

Section 1.3). After that, professional knowledge (see Section 1.3.1), diagnostic activities (see Section 1.3.2), and diagnostic accuracy (see Section 1.3.3), the three components of diagnostic competences, are described in more detail, and each is related to corresponding constructs used alternatively in teacher expertise research. This compilation results in further research gaps concerning the understanding and transferability of different constructs also included in these sections. An overview illustrates the variety of different constructs and conceptualizations used to investigate situation-specific skills in the context of assessment and diagnosis, as well as findings regarding their relationships to other components of diagnostic competences such as professional knowledge or diagnostic accuracy (see Section 1.4). As a last theoretical aspect, the use of videos in teacher education is considered (see Section 1.5), before the introduction closes with a description of the video-based assessment tool DiKoBi *Assess* (see Section 1.6). From the theoretical introduction, three research aims are derived (see Section 2), which were addressed in the context of three publications and two manuscripts (see Section 3). Afterward, the results of these publications are summarized and discussed with regard to the three research aims (see Sections 4.1 to 4.4). After referring to the limitations of the different studies conducted (see Section 4.5), further research is described that (can) follow the results of this dissertation (see Section 4.6). Finally, implications for research and teacher education are outlined (see Section 4.7).

Since several terms and constructs are considered relevant to understand the results of Section 3, Figure 1 was created to give an overview of the main terms and constructs used in this dissertation as well as the corresponding theoretical sections.

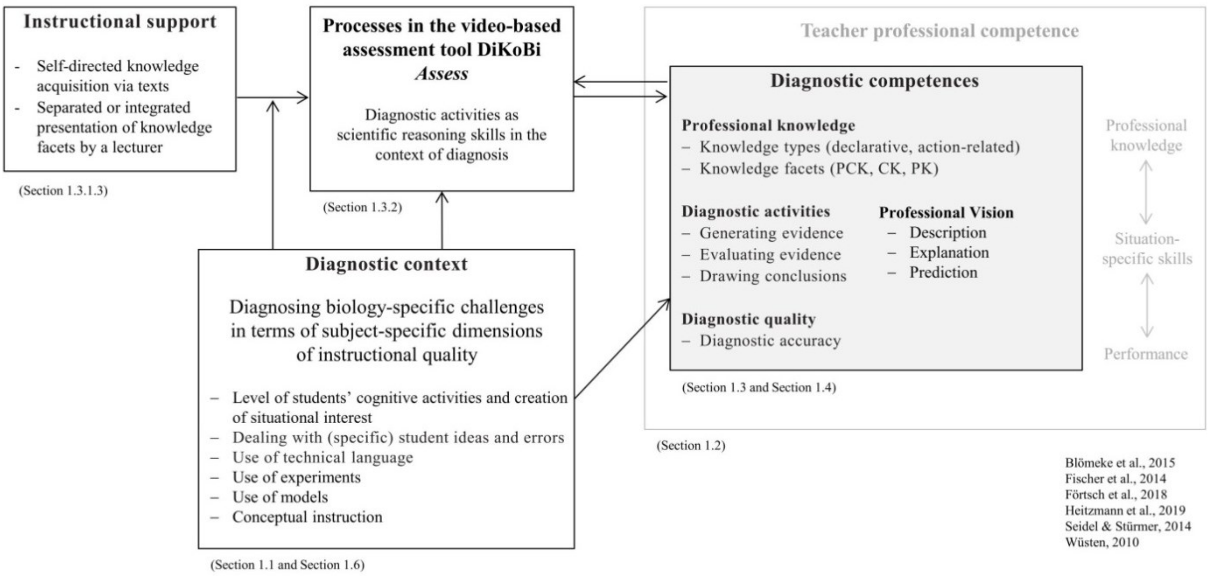

Figure 1. Overview of the main terms and constructs used in this dissertation. Diagnostic competences can be understood as a specification of teacher professional competence with regard to the context of diagnosis.

1.1 Instructional quality in biology classrooms

According to the process-mediation-product paradigm, the teaching conditions that are mostly set by the instructional behavior of a teacher affect student outcome variables such as student achievement (Brophy, 2000; Neuhaus, 2007; Seidel & Shavelson, 2007). A teacher's instructional behavior should aim to provide learning opportunities for students. Effective instructional characteristics that promote learning are also referred to as instructional quality features, which can be distinguished in generic and subject-specific features of instructional quality (Neuhaus, 2007; Seidel & Shavelson, 2007; Wüsten, 2010).

In German-speaking countries, generic features of instructional quality have been described by the use of three basic dimensions that were evolved based on the analysis of mathematics instruction but occur more or less across domains (Baumert et al., 2010; Klieme et al., 2001, 2006). The three basic dimensions are *classroom management*, *supportive climate*[3], and *cognitive activation*. Similar approaches can also be found in English-speaking countries. The CLASS-S framework (Pianta et al., 2012), for example, distinguished three broad domains of classroom interaction between students and teachers that are assumed to be important for student learning. The three domains, *classroom organization*, *emotional support*, and *instructional support,* can be seen as consistent with the three basic dimensions described in German-speaking research (Hamre et al., 2007).

Classroom management describes how teachers structure, organize, and monitor classroom instruction, and how they deal with student behavior. Overall, effective *classroom management* aims at generating time for students' learning activities (Klieme & Rakoczy, 2008; Kunter et al., 2007; Praetorius et al., 2018; Seidel & Shavelson, 2007). *Supportive climate* refers to characteristics on how a teacher addresses individual needs of students, which includes strategies of individualization and differentiation, but also adaptive support and feedback by the teacher, as well as positive teacher-student and student-student relationships (Hamre et al., 2007; Lipowsky et al., 2009; Praetorius et al., 2018). The third basic dimension, *cognitive activation*, requires instruction that promotes in-depth learning and reflection. The implementation of *cognitive activation* is to a great extent subject-specific, with good *classroom management* and *supportive climate* being conducive to the successful

[3] *Supportive climate* represents one out of several labels researchers used to name this basic dimension. While *supportive climate* has been used, for example, by Dorfner, Förtsch, and Neuhaus (2018) or Lipowsky et al. (2009), other researchers used labels such as *student orientation* (Klieme et al., 2001), *personal learning support* (Schlesinger & Jentsch, 2016), *student support* (Praetorius et al., 2018), or *emotional support* (Hamre et al., 2007).

application of cognitively activating strategies (Dorfner, Förtsch, & Neuhaus, 2018; Praetorius et al., 2018; Schlesinger & Jentsch, 2016). Key elements of cognitively activating instruction that foster conceptual understanding are the level of students' cognitive activities, conceptual instruction, and thoughtful discourse (Lipowsky et al., 2009). From a science education perspective, the level of students' cognitive activities is demonstrated by subject-specific features such as use of cognitive conflicts, focus questions, the activation of prior knowledge and conceptions (without targeting a specific answer), and the use of challenging and complex tasks to foster high-level thinking. These features indicate cognitively demanding instruction that foster students' deep understanding of the content (Förtsch et al., 2016; Förtsch et al., 2017; Lipowsky et al., 2009; Nawani et al., 2017; Praetorius et al., 2018). Subject specificity is then reflected in the fact that these features must be described with regard to the subject content that is taught, for example, in biology lessons (Dorfner et al., 2017; Neuhaus, 2007). The second key element of *cognitive activation*, conceptual instruction, aims at high knowledge linking through teaching interrelated facts and concepts instead of isolated facts, thus, fostering students' *cognitive activation* and the development of knowledge that is suitable for application (Brophy, 2000; Lipowsky et al., 2009). To support students' conceptual understanding, their prior knowledge and ideas have to be activated and linked to concepts they already know. Furthermore, learning opportunities should be provided, which allow students to apply and transfer the underlying concepts to different contexts (Brophy, 2000; Förtsch et al., 2020; Greeno, 2006; Wadouh et al., 2014). Thoughtful discourse features a teachers' questioning behavior that stimulates students to develop ideas in a reasonable and sustained manner, "to process and reflect on content, recognize relationships among and implications of its key ideas, think critically about it, and use it in problem-solving, decision making or other higher-order applications" (Brophy, 2000, p. 19). There is evidence that including these key elements of *cognitive activation* into mathematics or biology instruction increases instructional quality, and thus, student achievement (Förtsch et al., 2017; Lipowsky et al., 2009).

While it is undisputed that the implementation of the three basic dimensions is necessary for high-quality instruction, researchers also point out that "[o]ver and beyond these generic characteristics, content-related aspects of teaching quality need to be considered" (Praetorius et al., 2018, p. 423). In line with this claim to extend the framework of the three basic dimensions, research in recent years has discussed other characteristics that need to be implemented in a subject-specific way or are entirely specific to biology instruction and difficult to transfer to other subjects (Dorfner et al., 2017; Neuhaus, 2007; Wüsten, 2010).

Seidel and Shavelson (2007) showed that subject-specific characteristics are the ones that have the highest effects on student achievement. For biology instruction, several subject-specific dimensions and corresponding features have been described to extend the basic dimensions of instructional quality. These features were found to be empirically effective and can be implemented in the course of a lesson (Dorfner, Förtsch, Spangler et al., 2019; Wüsten, 2010). Table 1 presents subject-specific dimensions and selected features that serve as indicators of instructional quality in the biology classroom.

Table 1. Overview of subject-specific characteristics of instructional quality; modified and extended to Kramer et al. (2020).

Subject-specific characteristics of instructional quality		
Subject-specific dimensions	*Included subject-specific features (selection)*	*Reference/empirical evidence (selection)*
Cognitive activation		
a) level of students' cognitive activities	- challenging tasks and questions to foster high-level thinking - provoking cognitive conflicts - activating students' prior knowledge and conceptions - …	Brophy, 2000; Förtsch et al., 2017; Klieme et al., 2006; Lipowsky et al., 2009; Praetorius et al., 2018
b) conceptual instruction	- teaching interrelated facts and concepts to promote knowledge-linking - linking students' ideas, experiences, and prior knowledge with new concepts - encouraging students to discover and understand the meaning of underlying principles - …	Förtsch et al., 2020; Lipowsky et al., 2009; Wadouh et al., 2014
c) thoughtful discourse	- stimulating students to develop ideas thoughtfully and sustained - asking open-ended questions - …	Brophy, 2000; Lipowsky et al., 2009
Creation of content-related situational interest	- motivating presentation of the specific topic to create focused attention and affective involvement - referring to biology-specific everyday examples and problems - …	Dorfner, Förtsch, & Neuhaus, 2018; Klieme et al., 2006; Schiefele, 2009
Dealing with (specific) student ideas and errors	- eliciting ways of student thinking - detecting technical errors and using these errors as an opportunity for student learning (in terms of conceptual change) - encouraging students to explain their ideas and solution methods - …	Herppich et al., 2013; Praetorius et al., 2018; Rach et al., 2013; Spychiger, 2008
Use of technical language	- using minimally necessary topic-relevant terms - explaining and understanding new concepts in everyday terms before using scientific language - differentiate relevant terms clearly - …	Brown & Ryoo, 2008; Dorfner, Förtsch, & Neuhaus, 2019; Steffensky et al., 2015

Use of experiments	- using scientific inquiry methods to solve problems - steps of scientific inquiry are embedded: formulating scientific research questions and hypotheses, planning experiments, analyzing data, and drawing conclusions - …	Dorfner, Förtsch, Germ et al., 2018; Mayer, 2007; Simon, 1992; Tesch & Duit, 2004
Use of models	- model use goes beyond pure illustration - elaborate use of models (e.g., by using models as tools for scientific reasoning) - critical reflection of the model as a representation of something - …	Oh & Oh, 2011; Werner et al., 2017

Note: The features listed represent a selection of meaningful features. However, project-specific conceptualizations may draw on further indicators. In addition, *cognitive activation* was included as subject-specific part of the basic dimensions to illustrate that its implementation is to a great extent subject-specific.

For the implementation of generic and subject-specific instructional quality features, teachers' professional knowledge is considered crucial. Teachers' professional knowledge is seen as the essential basis that informs their actions (Krauss et al., 2004). For the instructional implementation of subject-specific features, teachers are expected to draw on the subject-specific facets of their professional knowledge: pedagogical content knowledge and content knowledge (Blömeke & Kaiser, 2017; Bromme, 1995; Jentsch et al., 2020; Wüsten, 2010). This knowledge, on the other hand, is not required when generic instructional features are implemented (Voss & Kunter, 2013).

1.2 Teachers' professional competence

Teachers‘ professionalism has long been studied from a cognitive perspective based on the expert paradigm (Bromme, 1997), focusing dominantly on knowledge facets of teachers. In the COACTIV project, for example, Baumert and Kunter (2013a) introduced a model of teachers' professional competence, in which cognitive aspects (professional knowledge) and non-cognitive aspects (beliefs, motivational orientations, self-regulation) are described. However, they defined professional knowledge as the core of teachers' professionalism being relevant for instructional practice, and therefore, for student achievement (Kunter & Baumert, 2013). Other empirical studies in mathematics or science used similar approaches to examine teachers professional competence, partly including motivational, metacognitive, or self-regulatory aspects (e.g., KiL: Großschedl, Harms et al., 2014; MT21: Schmidt et al., 2007; ProwiN: Tepner et al., 2012; QuIP: Fischer, H. E. et al., 2014; TEDS-M: Blömeke et al., 2010). However, since teaching is assumed to be multi-dimensional, the situated context in which knowledge is enacted needs to be considered as well (Blömeke & Kaiser, 2017; Depaepe et al., 2013). Investigating teachers' professional competence should therefore combine the cognitive perspective and the situated perspective to describe competence (Depaepe et al., 2013). As a result, Blömeke et al. (2015) modeled teachers' professional competence on a continuum that starts from dispositional traits (cognitive, affective, motivational) underlying teachers' cognitive skills. These relate to specific instructional situations and, in turn, lead to observable behavior (e.g., classroom instruction). In addition, teachers' instructional behavior is assumed to influence their cognitive skills, and thus, their dispositions as well (Carlson et al., 2019; Santagata & Yeh, 2016). Since the application of cognitive skills is considered to be situation-specific, cognitive skills are also referred to as situation-specific skills, whose conceptualization may vary with respect to the domain and the context (Stahnke et al., 2016).

The general conception of the competence as a continuum model by Blömeke et al. (2015) allows its transfer to specific contexts in which teachers have to act professionally. With regard to the importance of instructional quality features for effective teaching (see Section 1.1), teachers' ability to assess instruction in terms of the instructional quality features can be considered as one key aspect of teaching expertise that is a part of teachers' diagnostic competences (Artelt & Gräsel, 2009; Helmke & Schrader, 1987; Karst et al., 2014; Weinert et al., 1990). The importance of diagnostic competences is also acknowledged politically as an important element in teacher education. Therefore, it is stated in German standards of teacher education (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2008).

1.3 Diagnostic competences

Diagnostic competences and diagnosis are often related to medicine when a physician diagnoses health limitations of a patient or reasons for these limitations (Berner & Graber, 2008; Schwartz & Elstein, 2008). However, diagnostic competences are also necessary within classrooms and are largely used synonymously with assessment competences in German research (Herppich et al., 2018). Teachers' diagnoses in the classroom refer to the systematic assessment of a current state, either related to an individual student's learning process and achievement level or a whole classroom situation and the quality of instructional processes (Heitzmann et al., 2019; Helmke & Lenske, 2013; Klug et al., 2013). Based on these diagnoses, a teacher can adapt the learning environment to the learning requirements of the students by making decisions on how to continue with instruction in order to promote students' learning (Ostermann et al., 2019; Schrader, 2013). Therefore, teachers' diagnostic competences count as a crucial part of their expertise. At the same time, the processes described (systematically collecting, analyzing and interpreting data for assessment to inform pedagogical reasoning and decision-making) reveal similarities to scientific reasoning, from which the basic processes and skills used by a diagnosing person can be derived (cf. Brown et al., 2010; Wildgans-Lang et al., 2020). Therefore, some conceptualizations of diagnostic competences contain components that originate from research on scientific reasoning (e.g., epistemic activities that have been translated into diagnostic activities by Heitzmann et al., 2019; see Section 1.3.2). In the following, approaches to model diagnostic competences and corresponding components are briefly outlined.

In recent years, theoretical models of teachers' diagnostic competences have been developed in order to identify components of diagnostic competences, to understand interrelationships between them, and to investigate approaches to support them. For example, Klug et al. (2013) postulated a model of teachers' diagnostic processes when diagnosing students' learning behavior. The modeled diagnostic process consists of three phases (*preactional*, *actional*, *postactional*) that teachers go through during the diagnostic process. van Ophuysen and Behrmann (2015) described a model that differentiates quality features of the diagnostic process that are considered important for the diagnostic quality. The model distinguishes four components that teachers should use to systematically collect information about learners in order to justify instructional decisions. Two components refer to processes (data collection and processing of information), the other two components to the results (data and judgment), which have to be evaluated with respect to specific criteria. However, aspects such as knowledge, diagnostic occasions, or consequences of diagnostic judgments are often disregarded, although modeling diagnostic competences would, in fact, require a more comprehensive integration of these aspects (Praetorius & Südkamp, 2017).

A more comprehensive framework was recently introduced by Loibl et al. (2020) as the DiaCoM (*Explaining teachers' diagnostic judgements by cognitive modeling*) framework that distinguishes person characteristics (e.g., professional knowledge) and cognitive processes from observable situation characteristics (e.g., task difficulty, instructional quality features) and diagnostic behavior. Since the framework's purpose is to model diagnostic thinking, skills or activities that are utilized within diagnostic processes are not included. Additionally, the framework does not predefine how cognitive processes are modeled but emphasizes that this component needs to be specified with regard to a specific research goal. Other more comprehensive approaches defined diagnostic competences with regard to the competence as a continuum model as "those dispositions, situation-specific skills and performance that teachers need for diagnosis in the context of teaching and learning" (Hoth et al., 2016, p. 43). In that vein, Herppich et al. (2018) introduced a model of teachers' assessment competence that related constituents of judgment accuracy with constituents of the diagnostic process under consideration of individual dispositions and the specific diagnostic situation. However, the model is intended to be used to assess learners and characteristics relevant to learning, but not the diagnosis of other situations such as classroom instruction (Herppich et al., 2017).

Another framework for exploring diagnostic competences was used in the COSIMA project, an interdisciplinary project combining research in medical and teacher education (Heitzmann et al., 2019; Chernikova et al., in press). The conceptual framework builds on the recent attempt to integrate components of diagnostic competences from the cognitive and situated perspective and factors that influence its development. Therein, diagnostic competences are conceptualized across professional domains and contain both outcome and process-related variables. Similar to the conception of professional competence by Blömeke et al. (2015), the authors 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" (Heitzmann et al., 2019, p. 5). Therefore, three components of diagnostic competences emerged as vital: professional knowledge, diagnostic activities, and the quality of the diagnosis that can be described through diagnostic accuracy. Based on theoretical and empirical findings on the development of expertise and skills, the conceptual framework highlights the importance of individual learning prerequisites such as prior knowledge as well as students' engagement in complex, practice-oriented situations for improving diagnostic competences (Heitzmann et al., 2019; Renkl & Atkinson, 2003; van Lehn, 1996). With regard to research on complex learning environments (e.g., Lazonder & Harmsten, 2016), the authors also assumed that the inclusion of instructional support is effective for learning (Chernikova et al., in press). Figure 2 shows the main elements of the conceptual framework.

Pending is the evaluation of the framework with regard to the conceptualization of diagnostic competences as well as factors such as ways of instructional support that influence diagnostic competences within a specific domain such as biology education. Such questions are to be addressed in the project by means of video-based tools that provide approximations of practice (Grossman et al., 2009) (see Section 1.4). This dissertation takes on such questions to examine the model. In the following section, the three components are described in more detail.

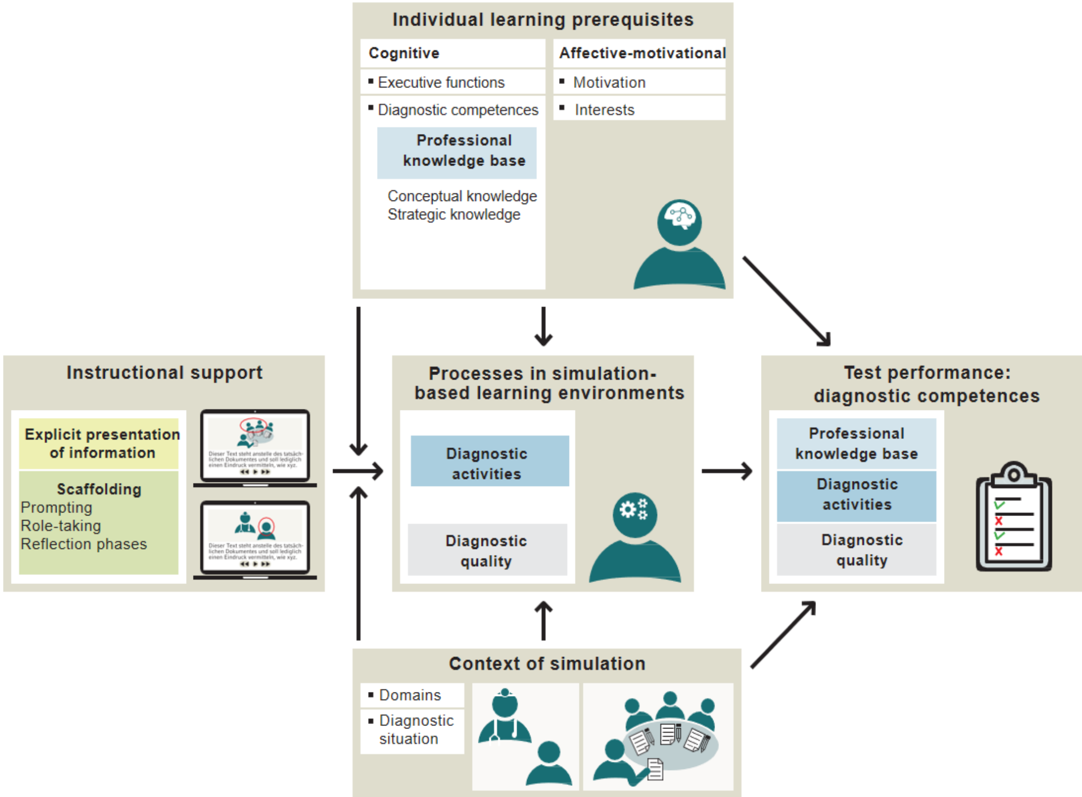

Figure 2. Illustration of the "COSIMA framework model" for using video-based tools to facilitate the development of diagnostic competences (https://www.for2385.uni-muenchen.de/aktuelles/rahmenmodellccby/cosima-frame-model_eng.pdf), created by COSIMA research unit, published in Chernikova et al. (in press), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/legalcode).

1.3.1 Professional knowledge base

Teachers' professional knowledge counts as an important part of teachers' diagnostic competences that enables them to apply specific skills such as diagnostic activities effectively (Blömeke et al., 2015; Brunner et al., 2013; Heitzmann et al., 2019). To systematically describe and analyze different aspects of the professional knowledge base, researchers distinguished different types and content facets of knowledge whose relevance may vary depending on the domain under study as well as the specific situation (Förtsch et al., 2018).

1.3.1.1 Facets of knowledge

On the basis of Shulman (1986, 1987), different facets of professional knowledge have been distinguished. Among them, three knowledge facets were relevant in a wide variety of studies: pedagogical-psychological knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK). Whereas PK can be seen as a content-independent facet, both CK and PCK are mainly considered content-specific (Baumert & Kunter, 2013a; Fischer et al., 2012; Jüttner et al., 2013).

The domain-general knowledge facet PK includes generic theories and methods of teaching and learning. One initial key aspect of PK covered knowledge about classroom management (König et al., 2011; Shulman, 1987). Conceptual frameworks of TEDS-M, COACTIV, and ProwiN additionally introduced general knowledge about instructional methods, performance assessment, and individual learning processes as further aspects (König et al., 2011; Lenske et al., 2016; Voss & Kunter, 2013). The aforementioned aspects also refer to contents related to the basic dimensions of instructional quality, which are also explicitly mentioned in some conceptualizations (Kunina-Habenicht et al., 2020).

CK refers to knowledge about facts and terms, as well as knowledge about subject-specific concepts and methods (Ball et al., 2008). Content knowledge also includes knowledge about connections between contents within one subject. A deep understanding means knowing many details about a specific content (Hill et al., 2005; Shulman, 1986). Therefore, CK is considered subject- and in parts content-specific (Hill et al., 2004).

The particular knowledge that is used to make particular content accessible for a particular group of learners is referred to as pedagogical content knowledge (PCK), containing aspects of content and pedagogy (Gess-Newsome, 2015; Shulman, 1987). For high-quality biology teaching, biology teachers need knowledge about students' (mis)conceptions regarding a specific instructional content as well as knowledge about subject-specific structures of instruction and corresponding teaching strategies (Depaepe et al., 2013; Park & Oliver, 2008; Schmelzing et al., 2013). Therefore, with regard to subject-specific instructional quality, biology teachers' subject-specific knowledge facets PCK and CK are of particular importance (e.g., Baumert & Kunter, 2013b). Given the many perspectives and situations under which PCK has been studied and the measuring methods that have been utilized (e.g., knowledge tests, interviews, assessments of lesson planning, classroom observations, or classroom reflections), many researchers argued for a more differentiated view on PCK, as the different situations and methods also resulted in different manifestations of PCK being captured, some of which were more closely to a skill than to a cognitive disposition (Alonzo & Kim, 2016; Gess-Newsome et al., 2019). As a result, science education researchers conceptualized a model that illustrates different forms or realms of PCK. The model is referred to as the Refined Consensus Model (RCM) and differentiates PCK into *collective* PCK (cPCK), *personal* PCK (pPCK), and *enacted* PCK (ePCK) (Carlson et al., 2019). While pPCK and ePCK are based on an individual's teaching and learning experiences, cPCK describes the knowledge that is held collectively by professionals within a domain. It is primarily informed from research (e.g., on teaching effectiveness), collated in subject-specific literature, and taught in subject courses (Carlson et al., 2019). Furthermore, cPCK provides the basis for the development of test instruments (cf. Sorge, Stender et al., 2019). pPCK develops due to a pre-service science teacher's education, practical experience, and professional exchange with colleagues. Therefore, pPCK reflects the individual structure and composition of content-specific pedagogical knowledge of a particular person that also informs the person's actions (Alonzo et al., 2019; Carlson et al., 2019). However, teachers also utilize PCK in the moment of teaching, when they plan instruction, carry out instruction, or reflect on instruction. Reflection may occur directly in action (e.g., when interacting with students), or as reflection on actions (e.g., when reflecting on classroom instruction) (Carlson et al., 2019). Researchers refer to this realm as ePCK. ePCK is considered to appear at a very short timescale and underly "science teachers' in-the-moment instructional decisions" (Alonzo et al., 2019, p. 274), thus, representing knowledge that is to a certain extent tacit and not articulable. In contrast, cPCK and pPCK represent rather static, articulable knowledge. Therefore, referring back to the different perspectives on professional knowledge, cPCK and pPCK may be important from the cognitive perspective, whereas ePCK builds a bridge to the situated perspective on competence, as it is generated in those situations where situation-specific skills are applied. Therefore, it may inform those skills that teachers apply during instruction and that have been operationalized as action-related skills (Kulgemeyer et al., 2020), professional vision (Seidel & Stürmer, 2014), or diagnostic activities (Heitzmann et al., 2019) (see Section 1.3.2). Finally, in the vein of the assumed bidirectional competence development, the generated ePCK can interact with other forms of PCK through reflection in and on action (Alonzo et al., 2019).

In summary, the differentiation of professional knowledge into facets and their further conceptualization is important to understand how knowledge and skills are connected and can be transformed in order to support the development of diagnostic competences.

1.3.1.2 Types of knowledge

In addition, the three professional knowledge facets can be distinguished into different cognitive types of knowledge (e.g., declarative, procedural, strategic, conditional). However, in different domains and research strands, their terminology partly differs with respect to the concrete understanding of each knowledge type. In order to compare research results from different domains, Förtsch et al. (2018) introduced a terminology to summarize emerged knowledge aspects. In summary, three knowledge types can be distinguished: knowing *that*, knowing *how*, and knowing *when and why*. Knowing *that* describes mostly declarative knowledge, meaning that someone knows about facts, terms, and principles of a specific subject and topic (Anderson, 1982; Paris et al., 1983; Tepner et al., 2012). It is complemented by knowledge about *how* to execute various actions, subject-specific instructional practices and procedures, as well as knowledge about *when and why* to execute these actions. Knowing *when and why* therefore describes conditions under which instructional decisions and practices are appropriate, as well as reasons for carrying out these practices (Förtsch et al., 2018; Paris et al., 1983; Tepner et al., 2012). Since knowing *how* and knowing *when and why* both contain knowledge that professionals need to act in various situations, they are also referred to as action-related knowledge types that underlie action-related skills (Förtsch et al., 2018; Kopp et al., 2008; Kulgemeyer et al., 2020). Action-related knowledge is also discussed under the terms procedural, conditional, or strategic knowledge. Still, all conceptualizations cover knowing *how* and knowing *when and why* as types of action-related knowledge, whose application is tied to a specific situation and context (Jüttner et al., 2013; Tepner et al., 2012).

This differentiation also has consequences for the measurement of the different knowledge types. Knowing *that*, for example, has mostly been measured through paper-pencil tests, requiring participants to remember and retrieve information (Großschedl et al., 2019; Schmelzing et al., 2013). Knowing *how* and knowing *when and why*, however, require assessment situations that provide practical contexts, such as utilizing text- or video vignettes in which authentic instructional situations are presented (Cauet et al., 2015; Hoth et al., 2018). In the field of biology education, current efforts are also being made to measure procedural knowledge about diagnostics and learning processes by using the simulated classroom biology SCR^{Bio} (Fischer et al., 2018). However, attempts exist, in which knowing *how* and knowing *when and why* have been successfully measured with paper-pencil tests that strove for more complexity (Blömeke et al., 2016; Jüttner et al., 2013). Still, such tests are scarcely used.

Overall, researchers agree that multiple methods are necessary to capture teachers' professional knowledge more comprehensively and that the understanding of the structure and nature of professional knowledge is essential to develop specific ways that promote knowledge development (e.g., Alonzo & Kim, 2016; Förtsch et al., 2018; Kleickmann et al., 2017; Lehane & Bertram, 2016).

1.3.1.3 Fostering knowledge development in teacher education

With regard to subject-specific characteristics of instructional quality in mathematics or biology classrooms, teachers' subject-specific knowledge facets and, in particular, their PCK are highly relevant (Förtsch et al., 2016; Kunter et al., 2013). Thus, one main research focus lies on the conditions for the development of PCK and the role of CK and PK (Kleickmann et al., 2017).

There is evidence that neither PK nor CK is sufficient for effective instruction and that PCK counts as an own unique knowledge facet with high relevance for subject-specific instructional quality (Baumert et al., 2010; Evens et al., 2018; Förtsch et al., 2016; Kleickmann et al., 2017). Concerning the subject-specific knowledge facets, research findings of different studies indicated a close relationship between CK and PCK, and that CK can be considered as relevant but not sufficient for the development of PCK (Baumert & Kunter, 2013b; Fischer et al., 2012; Kleickmann et al., 2017). Positive effects in terms of PCK development have been found for the explicit instruction on PCK that in turn affected PK and CK, as well as for a combined instruction of CK and PCK (Kleickmann et al., 2017; Tröbst et al., 2019). Instruction on PK, however, is primarily assumed to be important for providing domain-general knowledge about teaching and learning (e.g., about generic characteristics of instructional quality, see Wüsten, 2010). Holding that knowledge "is seen as a general precondition for a high quality of instruction […] [and thus] a necessary but not sufficient precondition to use CK and PCK for enhancing subject-specific learning processes" (Fischer et al., 2012, 443f.).

The assumed division into the three knowledge facets is also reflected in the tripartite organization of university teacher education, in which PK, CK, and PCK are taught in specialized faculties (König et al., 2018; Tröbst et al., 2019). Research findings from recent years, however, implicate benefits of an integrated presentation of knowledge facets that takes the existing relationship between the knowledge facets into account (Harr et al., 2014; Tröbst et al., 2019). Integration means addressing aspects of PK, CK, and PCK in an interrelated way that allows for direct cross-referencing between content and pedagogy, filling generic principles with subject-specific content and linking new and existing ideas, and thus, fostering applicable and transferable knowledge instead of inert knowledge (Evens et al., 2018; Harr et al., 2014; Renkl et al., 1996; Scott et al., 2011; Wadouh et al., 2014). Within teacher education, pre-service teachers acquire knowledge through various routes, for example, self-directed through texts, presented by a lecturer, or through practical approaches. However, an integrated treatment of the three knowledge facets PK, CK, and PCK is largely missing in each of these routes (Ball, 2000; Harr et al., 2014).

To date, integrated approaches have been examined in the context of computer-based intervention studies (Evens et al., 2018; Harr et al., 2014, 2015; Janssen & Lazonder, 2016). For example, Janssen and Lazonder (2016) were able to show that pre-service teachers not only preferred integrated instruction (in terms of content and pedagogy) but that integrated instruction of content and pedagogy was also more effective for the quality of PCK in lesson plans and PCK-related justifications. However, PCK was merely investigated as an outcome variable but not part of the integrated instruction. The high effectiveness of an integrated learning environment was also shown by Harr et al. (2014, 2015). They found that integrating relevant aspects of PK and PCK increased the applicability of PK aspects and the simultaneous application of PK and PCK when participants had to solve a particular case in classroom teaching. Considering the effort it would take to restructure curricula at universities, Harr et al. (2015) pursued a different approach. In this approach, integration was not provided by a pre-structured environment but had to be done by the pre-service teachers themselves. Pre-service teachers were mentally triggered by specific prompts to relate content topics and pedagogical principles. Due to the effectiveness of these prompts, Harr et al. emphasized the high potential of prompted integration for implementability in teacher education—albeit at the expense of time. Contrary to these promising results of integrated instruction, Evens et al. (2018) found no differences between integrated and separated instruction on the three knowledge facets on the development of PCK in the domain French as a foreign language. The authors trace this back to participants being overwhelmed by the integrated information, the utilized instruments that did not account for integrated knowledge, or their study design with particular focusing on teachers' PCK.

However, none of the aforementioned studies addressed all three knowledge facets in terms of instruction on a science subject such as biology.

1.3.2 Diagnostic activities: situation-specific skills for diagnosis

Teachers' professional knowledge underlies the application of their situation-specific skills that can be described as diagnostic skills within diagnostic contexts (Brunner et al., 2013; Hoth et al., 2016). Since the assessment of diagnostic skills requires a situated approach, the use of video has become a useful tool to investigate diagnostic processes and corresponding situated skills (e.g., Blomberg et al., 2013; Kersting, 2008; Sherin & van Es, 2009) (see Section 1.5). However, skills have also been measured by creating lesson plans or by prompting teachers to explicitly explain their planning, enactment, and reflection on a lesson within interviews (Alonzo & Kim, 2016; Lee & Luft, 2008).

Different conceptual approaches exist to investigate situation-specific skills for diagnosis. They differentiate steps or phases of the diagnostic process (Herppich et al., 2018; Klug et al., 2013), describe different analytical or reflective abilities relevant for assessment situations (e.g., Benedict-Chambers & Aram, 2017; Seidel & Stürmer, 2014; van Es & Sherin, 2002) or connect to constructs such as formative assessment (e.g., Furtak et al., 2016). Theoretically, the models can be related to each other since the execution of specific process steps is based on the ability to execute the respective process steps. However, for relating models, corresponding skills need to be described. Klug et al. (2013) introduced three phases of the diagnostic process that teachers undertake cyclically: *preactional* (involves knowledge activation in order to set aims, and recall quality criteria for testing and judgment formation), *actional* (involves systematical enactment, for example, by collecting information, selecting and interpreting relevant information, making predictions), and *postactional* (involves the implementation of a pedagogical action that follows the diagnosis, for example, feedback). Similar to those phases, Herppich et al. (2018) modeled teachers' diagnostic process by several steps such as the formation of hypothesis or collection of information that are connected through different pathways a teacher can take. However, both the model by Klug et al. (2013) and the process model by Herppich et al. (2018) lack a precise description of skills relevant to succeed in diagnostic situations. Another approach to diagnostic skills builds on formative assessment practices. Researchers defined formative assessment as the continuous diagnosis "of the individual learning progress and the continuous response to promote learning" (Vogt & Schmiemann, 2020, p. 4). In that vein, formative assessment abilities such as designing tasks, asking questions, interpreting student ideas, or providing feedback might be understood as components of diagnostic competences that influence subject-specific instruction (cf. Furtak et al., 2016; Rakoczy et al., 2017). However, the concept of formative assessment is not defined precisely yet, nor does research explicitly connect teachers' professional knowledge to formative assessment practice (Herppich et al., 2018).

A more precise approach was taken by Heitzmann et al. (2019), who transferred scientific reasoning skills used for the evaluation of data (cf. Fischer, F. et al., 2014) to the diagnostic process. In that vein, the diagnostic process can be seen as an evidence-based scientific reasoning process on instruction that involves systematic observations, evidence generations and evaluations, and drawing conclusions (Brown et al., 2010). As a result of the transfer, eight diagnostic activities were described that teachers use to identify, interpret, and make diagnostic decisions based on noteworthy classroom events (see Table 2).

The researchers emphasized that there is no specific order of executing diagnostic activities, nor is there an obvious relationship between the individual processes (Heitzmann et al., 2019; Scheiner, 2016). Thus, the different activities can be seen as a repertoire from which appropriate activities are applied depending on the domain and the diagnostic situation. In the context of teacher education, the diagnostic activities *generating hypothesis*, *generating evidence*, *evaluating evidence*, and *drawing conclusions* were considered to be particularly relevant (Bauer et al., 2020; Wildgans-Lang et al., 2020). Overall, the diagnostic activities are considered to be relevant for generating knowledge that is the basis for diagnosing and making further decisions (Heitzmann et al., 2019).

Table 2. Taxonomy of the diagnostic activities according to Heitzmann et al. (2019) and Fischer, F. et al. (2014)

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 student's 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. They interpret the data, thus making sense of the generated evidence with regard to their belief, knowledge, and expertise (cf. Lai & Schildkamp, 2016).
Drawing conclusions	As a result of the evaluation of 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.

Furthermore, analytical abilities for the assessment of classroom situations have been studied using different conceptual approaches that have been labeled as PID (including perceptual, interpretative, and decision-making skills, cf. Kaiser et al., 2015), teacher noticing (van Es & Sherin, 2002), professional vision (Seidel & Stürmer, 2014), or reflective practice (Benedict-Chambers & Aram, 2017; Hiebert et al., 2007). Despite differences in the specific definitions, the conceptual approaches show considerable overlap, with a predominant focus on selective attention and knowledge-based reasoning (Gaudin & Chaliès, 2015). Both of these cognitive processes have been described in detail in the professional vision approach. *Professional vision* was first described by Goodwin (1994, p. 606) as "socially organized ways of seeing and understanding events that are answerable to the distinctive interests of a particular social group.” In the context of teaching, *professional vision* can be understood as the ability to notice and reason about relevant features of classroom situations based on the quality of teachers’ professional knowledge (Seidel & Stürmer, 2014; van Es & Sherin, 2002). *Noticing* that is also termed perception or selective attention refers to whether teachers pay attention to events in the classroom that are relevant to learning and corresponding instructional quality features as described by teaching effectiveness research (Kaiser et al., 2015; Seidel & Stürmer, 2014; Sherin & van Es, 2009), for example, perceiving deficient student ideas or identifying obstacles to learning in order to even use them for learning if necessary. *Reasoning* refers to sense-making processes of what has been noticed in the classroom. Three main aspects of these processes have been distinguished: *description*, *explanation*, and *prediction*. In that vein, reasoning can be assumed as follows: First, a noticed event is described without any judgments. Based on teachers’ professional knowledge, the event is then linked with appropriate concepts and principles that support its relevance in terms of student learning and teaching effectiveness, thus, also justifying the interpretive stance taken (Seidel & Stürmer, 2014; van Es & Sherin, 2002). And third, possible consequences in terms of student learning are predicted in order to inform subsequent pedagogical decisions (Seidel & Stürmer, 2014; van Es & Sherin, 2002). These pedagogical decisions usually follow professional *reasoning* and can comprise the anticipation of responses to students’ answers or activities, or the proposal of alternative teaching strategies (Kaiser et al., 2015; Santagata & Yeh, 2016). However, while *professional vision* is often studied in the context of video-based research, only a few studies have used videos to promote decision-making skills (Santagata et al., 2021).

In the vein of the competence as a continuum model, teachers’ situation-specific skills in terms of *noticing* and *reasoning* are considered to be knowledge-based since “knowledge guides attention processes and reasoning about observed events” (Schäfer & Seidel, 2015, p. 38). Since *noticing* and *reasoning* show similarities to classroom diagnosis and since the aspects *description*, *explanation*, and *prediction* described therein may also be considered as activities, the question arises to which extend conceptualizations of situation-specific skills such as *diagnostic activities* and cognitive-psychological constructs such as *professional vision* are comparable or whether they can be transferred (Scheiner, 2016; van Ophuysen & Behrmann, 2015).

1.3.3 Diagnostic quality in terms of diagnostic accuracy

The quality of the diagnosis must also be considered an important component of diagnostic competences that has predominantly been studied in research on teachers' judgment accuracy (Artelt & Rausch, 2014; Heitzmann et al., 2019; Herppich et al., 2018; Südkamp et al., 2012). However, there are approaches in which judgment accuracy has been used as the only measure for operationalizing diagnostic competences, although this approach has been criticized in recent years (e.g., Praetorius & Südkamp, 2017; Ufer & Leutner, 2017; van Ophuysen, 2010). With regard to the situated perspective on diagnostic competences, both the processes that lead to diagnostic judgments as well as the individual resources that allow these processes must also be taken into account in order to support pre-service teachers in developing their competences (Herppich et al., 2018). Therefore, diagnostic accuracy should be considered as one component of diagnostic competences that need to be complemented by other components such as teachers' professional knowledge and their diagnostic skills (Artelt & Rausch, 2014; Heitzmann et al., 2019).

In general, the accuracy of teachers' diagnoses can be assessed by the difference between the teacher's individual judgment and more objective assessments of performance, that can be, for example, an expert's diagnosis or the comparison with implemented instructional quality features described in the scientific literature (Artelt & Rausch, 2014; Kersting et al., 2012; Schäfer & Seidel, 2015). Judgment accuracy has been assessed in terms of diagnosing individuals' characteristics or task-related characteristics, for example, in the context of *the simulated classroom*, a computer simulation in which the user takes on the role of a teacher, interacts with virtual students, and assesses students' experimentally controlled performance (Südkamp et al., 2008). Accuracy indicators included, for example, the assessment of students' performance levels and their ranking (Artelt & Rausch, 2014; Karst et al., 2014; Südkamp et al., 2008), the diagnosis of misconceptions (Wildgans-Lang et al., 2020), or the selection of tasks with an appropriate level of difficulty (Karst et al., 2014; Kron et al., 2021; Schrader, 1989).

Regarding the broader understanding of the concept of diagnostic competences, teachers' ability to correctly diagnose classroom incidents is vital in order to adapt instruction effectively (Artelt & Rausch, 2014). Applying accuracy measures to teachers' diagnostic skills can therefore be seen as an important quality criterion that has already been discussed, for example, in terms of perception accuracy (Carter et al., 1988). Perception accuracy refers to the precise observation of classroom situations and has been studied in terms of expert-novice-comparisons (Carter et al., 1988; Kaiser et al., 2015). However, accuracy should also be considered for the application of other diagnostic skills, such as diagnostic activities, but this has not been investigated sufficiently yet. First attempts to differentiate diagnostic accuracy regarding the use of diagnostic activities on different knowledge facets exist, for example, in the context of *the student inventory*. *The student inventory* is a digital instrument to diagnose virtual student exams on the biological topic evolution (Fischer et al., 2021). Therein, diagnostic accuracy was assessed for the application of CK (i.e., diagnosis of scientific correctness), PCK (i.e., diagnosis of a specific misconception), and PK (i.e., diagnosis of achievement-relevant information). However, diagnostic activities themselves have not been further operationalized.

Overall, accuracy measures have intensively been studied with regard to indicators of student performance. However, analyses of teachers' diagnostic accuracy when diagnosing instructional features in situated contexts such as a science or biology classroom are missing. That may be due to the greater technical effort required when features of instructional quality need to be recorded in order to use them in studies on diagnostic competences (cf. Karst et al., 2014).

1.4 Relationships between components of diagnostic competences

Although studies on the components of diagnostic competences are available, they rarely consider all three components together. For example, although it is assumed that knowledge and specific skills play a role for accurate diagnoses, their interrelationship has not yet been systematically studied. In addition, there is a wide variety of conceptual approaches and constructs used to operationalize components of diagnostic or assessment competences, making a comparison or transfer of results difficult. Table 3 provides an overview of constructs that have been utilized to investigate situation-specific skills in the context of classroom diagnosis and relevant findings.

Overall, it becomes apparent that knowledge, skills, and accuracy measures are related, but depending on the context, different knowledge facets are of importance. The specific knowledge facets, in turn, predict certain skills. For subject-specific aspects of diagnosis, content-related facets in terms of CK and especially PCK are considered decisive (e.g., Hoth et al., 2016; Kron et al., 2021; Meschede et al., 2017), while for diagnosing general pedagogical aspects, PK is important (e.g., König & Kramer, 2016). However, the studies also showed that, in most cases, only individual components of diagnostic competences were investigated.

Table 3. Overview of research findings about the relationships between the three components of diagnostic competences according to the definition of the COSIMA framework. However, besides the constructs used therein, many other constructs have been used that can be parallelized for the most part. The selected constructs can be considered relevant with respect to diagnostic contexts. The findings listed for the respective constructs represent a selection of studies that investigated at least two of the components related to diagnostic competences.

Theoretical construct	Sample and context	Professional knowledge	Diagnostic activities	Diagnostic accuracy	Findings
Diagnostic activities	– primary school pre-service teachers of mathematics – diagnosing student's misconceptions and competence levels in mathematics		x	x	Wildgans-Lang et al. (2020): – quality of diagnostic activities was more important for diagnostic accuracy than the frequency of diagnostic activities – the accurate diagnosis of misconceptions was found to be more challenging than the diagnosis of competence levels
Diagnostic skills	– secondary school in-service teachers of geography, biology, and German – text-picture integration	x		x	McElvany et al. (2009): – only poor correlations between PCK or the duration of professional experience and diagnostic accuracy
Judgment accuracy	– secondary school pre-service teachers of mathematics – diagnosing task requirements	x		x	Ostermann et al. (2018): – Instructional input on students' misconceptions improved judgment accuracy (ranking tasks according to their difficulty)
Process model of teachers' diagnostic competence	– secondary school pre- and in-service teachers (subject not specified) – diagnosing students' learning behavior	x	(x) (x)	 x	Klug et al. (2016): – high knowledge of diagnostics predicted pre-service teachers diagnostic competence regarding the three dimensions preactional, actional, and postactional Klug et al. (2013): – preactional, actional, and postactional dimension scores predicted diagnostic accuracy (16 % of explained variance with the actional dimension explaining the greatest part)
Assessment competence	– secondary school pre-service teachers of languages, natural or sport sciences – assessment of learning strategies	x		x*	Glogger-Frey et al. (2018): – knowledge (in pieces) of comprehension-oriented learning strategies predicted the quality of the assessment performance of learning strategies

Theoretical construct	Sample and context	Professional knowledge	Diagnostic activities	Diagnostic accuracy	Findings
PID (perceiving, interpreting, decision-making)	– secondary school in-service teachers of mathematics – mathematics instruction and general classroom management	x	x*	x	Blömeke et al. (2016): – teachers prepared to teach both upper-secondary and lower-secondary mathematics had stronger prerequisites and much more mathematics-related learning opportunities during teacher education that resulted in a stronger cognitive base (math CK, math PCK, and a measure of accuracy under time pressure) – their cognitive base was better connected to their situation-specific cognition (including skills [PID] and general PK measured with a strong practical focus) – stronger situation-specific skills were related to classroom performance of higher quality (indicated by grades in the practical teaching exams)
Noticing linked to PID	– primary school in-service teachers of mathematics – mathematics instruction and student learning	x	(x)		Hoth et al. (2016): – defined PID as skills fo situation-based diagnostic competence – mathematical knowledge was connected with a content-related perspective (visible in the application of skills) in situation-based diagnoses – higher general pedagogical knowledge was connected to a stronger focus on classroom management, organizational aspects and other pedagogical facets
	– primary school in-service teachers of mathematics – pedagogical and didactical aspects in mathematics instruction	x	(x) (skills were executed by participants, but the focus of analysis was on the diagnostic result)		Hoth et al. (2018): – teachers with high professional knowledge (both subject-specific and general pedagogical knowledge) noticed crucial elements of student learning – teachers with low professional knowledge focused on elements that were not relevant for understanding a student's error (and thus not accurate)
Noticing and Interpreting	– secondary school pre-service teachers of mathematics – pedagogical situations in a mathematics classroom	x	x*		König et al. (2014): – declarative-conceptual general PK did not correlate significantly with noticing, but general PK correlated significantly with interpreting
	– secondary school pre- and in-service teachers (subject not specified) – classroom instruction referring to classroom management situations	x	(x) (joint measure including aspects of skills and accuracy)		König and Kramer (2016): – general PK correlated moderatly with classroom management expertise (that is formed by three cognitive demands: accuracy of perception, holistic perception, and interpretation/justification of action) – classroom management expertise was more highly intercorrelated with the PK-dimension *classroom management/motivation* than with any other PK-dimension – classroom management expertise seems more dependent on classroom expertise, whereas PK can be acquired during theoretical teacher education at university

Theoretical construct	Sample and context	Professional knowledge	Diagnostic activities	Diagnostic accuracy	Findings
Professional vision	– primary school pre- and in-service teachers of science (*Sachunterricht*) – classroom management and learning support	(x)	x*		Steffensky et al. (2015): – pre-service and in-service teachers with specific science-related education had higher scores for professional vision of learning support (that is subject-specific)
	– primary school pre- and in-service teachers of science (*Sachunterricht*) – cognitive activation and structuring in elementary science education	x	x*		Meschede et al. (2017): – PCK correlated moderately with professional vision (noticing and interpreting) – practical experience might have an impact as well
	– pre-service teachers (school type and subject not specified) – pedagogical aspects of effective teaching and learning	x	x*		Stürmer et al. (2012): – video-based course (that focused on observing and interpreting components of effective teaching in authentic classrooms) promoted the acquisition of pre-service teachers' declarative knowledge most – participation in video-based course (compared to self-regulatory learning course and regular teaching of general aspects of teaching and learning course) led to higher gains in prediction ability (pre-service teacher could better transfer pedagogical concepts to authentic classroom situations)
Classroom video analysis assessment (CVA)	– secondary school in-service teachers of mathematics – mathematics instruction	x	x*		Kersting et al. (2012): - teachers' mathematical knowledge for teaching (including CK and PCK) strongly correlated with their ability to analyse video clips of classroom interactions - teachers' ability to analyse video clips predicted instructional quality
Formative assessment	– secondary school in-service teachers of biology – learning progressions for natural selection	(x)	x*		Furtak et al. (2016) and Furtak et al. (2018): – indications that teacher knowledge (acquired through the course of a study) may have impacted formative assessment abilities (higher-quality questions and higher-quality feedback) – however, knowledge was not measured directly

Note: The table also provides an overview of the components of diagnostic competences that are directly or indirectly addressed in the studies. "x" refers to a direct and explicit study of the component. "(x)" means that the respective component was not measured directly but was inferred by indicators. "x*" indicates an explicit measure but that a conceptualization similar to diagnostic activities/diagnostic accuracy was used for operationalization. However, these conceptualizations are not congruent with the components of diagnostic competences.

1.5 Use of videos in teacher education

For pre-service teachers, the development of a well-founded knowledge base, the ability to apply this knowledge in classroom situations relevant for teaching and learning, and effective instructional decision-making is challenging (Kaiser et al., 2017; Star & Strickland, 2008). In the scientific discourse on the promotion of professional competence, the use of videos is highly emphasized, since videos are considered promising to approach real-life demands (Kersting, 2008; Seidel & Stürmer, 2014; Star & Strickland, 2008). Moreover, previous research results indicate the effectiveness of video-based professional development programs with regard to the promotion of pre-service teachers' situation-specific skills (e.g., Sherin & van Es, 2009; Stahnke et al., 2016).

Currently, videos are being used in a variety of situations and for a variety of purposes in teacher education (Blomberg et al., 2013; Darling-Hammond, 2010). Video-based programs and tools have been developed that differ, for example, in:

- video type (e.g., whole lesson or video clip, real-life or scripted classroom scenarios),
- video object (e.g., reflecting on one's own teaching or other teacher's practice),
- participation structure (e.g., view videos individually, in pairs, or in groups),
- video viewing support (e.g., no instruction or broad prompts, specific questions to be answered, assessment of rating scale items), and
- the way how they achieve learning goals (e.g., show characteristic professional situations, best-practice examples, evaluate competencies) (Gaudin & Chaliès, 2015; Santagata et al., 2021; Sherin & van Es, 2009).

Recently, the number of developed video-based tools has increased significantly (e.g., Gold & Holodynski, 2017; Kaiser et al., 2015; Kersting, 2008; König & Kramer, 2016; Meschede et al., 2017; Michalsky, 2014; Piwowar et al., 2018; Seidel et al., 2010; Wiedmann et al., 2019; Wiens et al., 2020)[4]. Among other things, the tools were designed for the assessment of situation-specific skills such as professional vision. Furthermore, videos have been used to create video-based simulations that can be understood as a specific type of video-based tools (e.g., Codreanu et al., 2020; Pickal et al., in press). A simulation represents a simplified but valid, true and dynamic model of reality implemented in a digital system (Sauvé et al., 2007). Thus, video-based simulations allow pre-service teachers to interact with the system, explore

[4] Note that the examples given can only be understood as a selection of tools relevant in the broader context of diagnostic competences, since further tools are continuously being developed, validated, and adapted to specific topics or domains.

it in an inquiry manner, and make decisions, which altogether is intended to enhance diagnostic competences (Codreanu et al., 2020; Heitzmann et al., 2019; Sauvé et al., 2007). Still, empirical evidence is scarce. However, simulations do not have to be video-based per se. There are also simulations in which texts and images are used to create virtual classroom situations (e.g., Südkamp et al., 2008), or in which live role-plays are implemented (e.g., Kron et al., 2021). In this dissertation, however, the focus is on video-based tools such as video-based simulations.

Overall, two overarching objectives of using videos in teacher education can be differentiated: (1) Within educational courses, the use of videos is considered beneficial for the promotion of knowledge and situation-specific skills such as professional vision or diagnostic activities necessary to interpret and reflect on instruction and to derive further instructional strategies (Behling et al., 2019; Gaudin & Chaliès, 2015). (2) Within situated research approaches, including the investigation of diagnostic competences, using videos is beneficial for measuring situation-specific skills such diagnostic activities since they have to be applied within very specific contexts, which closely resembles their use in real-life classroom situations. Videos thus serve as an approximation of practice (Grossman et al., 2009). In addition, they can break down complex instructional practice into smaller units, which is called decomposition of practice (Grossman et al., 2009). Decomposition of practice is considered to reduce the complexity of a specific diagnostic situation, for example, by preparing short classroom situations that include selected, noteworthy incidents in teaching and learning. This allows adapting to pre-service teachers' assessment skill level (Codreanu et al., 2020).

For the investigation of pre-service teachers' diagnostic competences, tools have been developed that contained not only real-life videos (e.g., Steffensky et al., 2015) but also scripted (staged) classroom scenarios, in which specific challenges relevant for effective instructional behavior within the science classroom, and thus, for pre-service teachers' learning, can be embedded and have been experienced as authentic (Codreanu et al., 2020; Piwowar et al., 2018). Similar to real-life videos, staged videos can activate pre-service teachers' professional knowledge and skills by the specific context shown in the videos (Hoth et al., 2018; Kersting, 2008; Seidel & Stürmer, 2014). When measuring professional knowledge and specific skills auch as diagnostic activities, open-ended items that require teachers to elaborate on their observations and interpretations (e.g., Kaiser et al., 2015; Kersting, 2008), or rating scale items that can be agreed or disagreed with have been used (e.g., Meschede et al., 2017; Seidel & Stürmer, 2014). However, researchers underlined higher face validity in terms of an open-ended approach, since the measurement of complex cognitive processes, such as those that occur during diagnosing, requires pre-service teachers to actually engage in them and make their perspective explicit in their own words (Arffman, 2016; Kaiser et al., 2015).

Moreover, staged videos can be designed regarding different foci. The interaction of a teacher with a group of students or individual students can be shown in the videos. But the observer can also monitor the entire class that is considered more complex (Kersting, 2008; König & Kramer, 2016; Tolsdorf & Markic, 2017). Attention may be paid to the instructional behavior of a teacher with regard to the treatment of student responses, but the focus may also be on instructional quality features that have been proven to impact student learning (cf. Kersting et al., 2012; Meschede et al., 2017; see Section 1.1). However, using videos to investigate pre-service teachers' diagnostic competences with regard to dimensions of instructional quality has only been considered in few approaches in science education yet (e.g., Meschede et al., 2017; Roth et al., 2011). Therefore, the range of contexts and content-specific domains addressed in video-based tools as well as addressed aspects of instructional quality need to be extended in the study of skills such as professional vision or diagnostic activities (Helmke & Schrader, 1987; Meschede et al., 2017; Sherin & van Es, 2009).

1.6 The video-based assessment tool DiKoBi *Assess*

Based on the importance of diagnostic competences for teachers and the potential to investigate them using video-based tools, the video-based tool DiKoBi *Assess* was developed in the COSIMA project. The implementation of the tool and its different versions in the online learning platform Unipark (Questback GmbH, 2018) was done to a large extent in the context of this dissertation.

Overall, three versions of DiKoBi *Assess* have been created (DiKoBi I *Assess*, DiKoBi II *Assess*, and DiKoBi III *Assess*), so that participants do not have to diagnose the same lesson if they use DiKoBi *Assess* more than once. All three versions cover the biological topic "skin", representing an important topic area within science curricula in different school types, grades, and nations (e.g., National Research Council, 2012; State Institute of School Quality and Educational Research, Munich, 2017). However, each version of DiKoBi *Assess* refers to a different subtopic. The subtopics are "skin as a sensory organ" (addressed in DiKoBi I *Assess*), "protective functions of the skin" (addressed in DiKoBi II *Assess*), and "importance of the skin for the regulation of the body temperature" (addressed in DiKoBi III *Assess*). However, in the studies conducted in the context of this dissertation, only the versions DiKoBi I *Assess* and DiKoBi II *Assess* have been utilized. Even though the different versions differ in terms of their content, the subject-specific dimensions and features of instructional quality addressed therein are the same as they can be applied to different contents. Since the structure of the tool is therefore the same in all versions, only the label DiKoBi *Assess* is used in the following description of the tool's structure.

The development of DiKoBi *Assess* was based on subject-specific instructional quality features that teachers should be able to identify and implement in biology instruction to improve student achievement. Based on a literature review and with reference to previous research results, several biology-specific dimensions of instructional quality were identified, which can be described by different features of instructional quality (see Table 1 in Section 1.1). Based on these dimensions, six classroom situations were scripted and videotaped, in which a teacher implemented corresponding instructional quality features in a deficient way or not at all. The missing features are therefore also referred to as subject-specific challenges that are to be diagnosed. Figure 3 gives an overview of the videotaped classroom situations, the instructional quality dimensions and the challenges included.

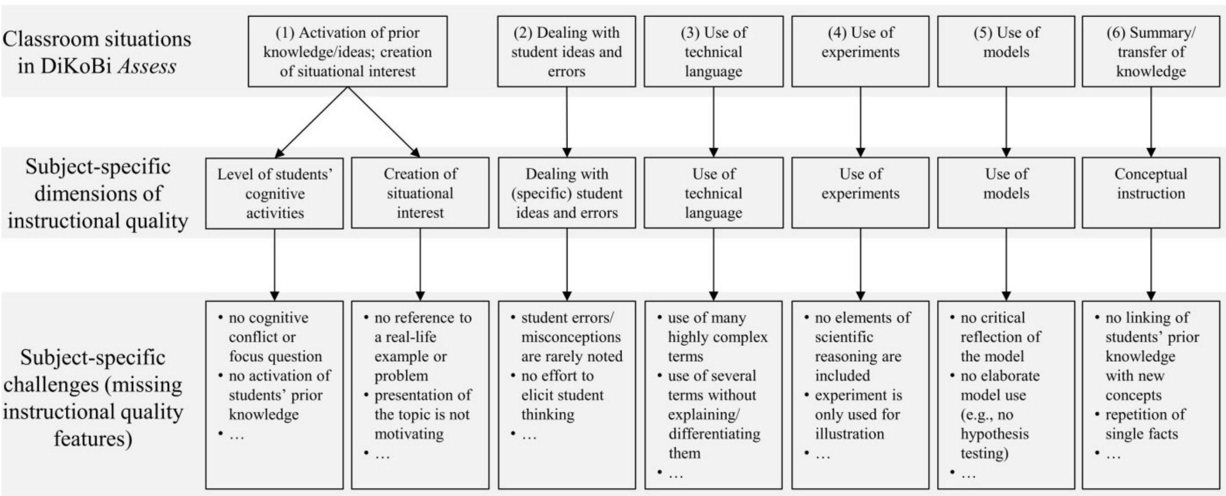

Figure 3. Overview of classroom situations embedded in DiKoBi *Assess*, their underlying subject-specific dimensions of instructional quality, and included subject-specific challenges that are based on specific instructional quality features. This structure is the same for all versions of DiKoBi *Assess*.

The six videotaped classroom situations are embedded in DiKoBi *Assess* in such a way that they follow the course of a lesson, with classroom situation (1) *Activation of prior knowledge/ideas and creation of situational interest* representing the beginning of the lesson and classroom situation (6) *summary/transfer of knowledge* representing the final transfer phase (cf. Dorfner, Förtsch, Spangler et al., 2019). The classroom situations (2), (3), (4), and

(5) represent different elaboration phases within one lesson. This simulation of the course of a lesson is intended to increase the authenticity of the diagnostic situation by simulating a real observation situation of biology instruction. However, each of the videotaped classroom situations represents a self-contained instructional situation that can be assessed independently of the other classroom situations as well.

Moreover, the diagnostic process is structured by three tasks that are the same for all classroom situations (see Appendix A for an example). Each of the three tasks intends to measure another diagnostic activity. Based on activities that have been highlighted and utilized in video-based research on diagnostic competences (see Section 1.3.2, cf. Bauer et al., 2020; Kaiser et al., 2015; Seidel & Stürmer, 2014), the diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions* are considered particularly relevant in the context of diagnosing instructional quality. Therefore, the three tasks are as follows: First, the videotaped classroom situation is to be observed, and subject-specific challenges of biology instruction are to be described (Task *Describe*). Therefore, Task *Describe* refers to the diagnostic activity *generating evidence*. Second, the relevance of the observations described has to be justified with reference to didactical theories or subject-specific concepts (Task *Explain*). Therefore, Task *Explain* refers to the diagnostic activity *evaluating evidence*. Third, based on the critical observations, alternative teaching strategies are to be proposed (Task *Alternative Strategy*). The third task, therefore, refers to the diagnostic activity *drawing conclusions*.

In addition, the data to be collected that refers to the diagnostic activity *generating evidence* will also be used to measure diagnostic accuracy. For a high accuracy measure, the corresponding biology-specific challenges of the respective classroom situations must be addressed in the participants' answers. This approach corresponds to the concept of perception accuracy as described in Section 1.3.3 (cf. Carter et al., 1988; Kaiser et al., 2015).

Overall, DiKoBi *Assess* is intended to measure components of diagnostic competences in terms of diagnostic accuracy and the diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions* within a situated practice-oriented context.

2. Aims

The video-based tool DiKoBi *Assess* serves as a core element for investigating the diagnostic competences of pre-service biology teachers. In DiKoBi *Assess*, the diagnostic focus lies on the instructional behavior of a biology teacher and the resulting instructional quality features. Previous tools lack such a diagnostic focus, although this focus is considered to be important for the implementation of diagnostic quality and student learning. Therefore, a need for research in this regard exist (Meschede et al., 2017). The validation of DiKoBi *Assess* is still pending. With regard to the COSIMA framework (see Section 1.3), the analyses in this dissertation will also provide data for the evaluation of the framework components *video-/ simulation-based tool*, *diagnostic competences*, and *instructional support* (Heitzmann et al., 2019; Chernikova et al., in press).

As the theoretical introduction has highlighted, there are various constructs that have been used to operationalize diagnostic competences and their components. In part, these constructs show striking similarities to more analytical constructs such as *professional vision*, which is well-established in video-based research on teachers' professional competence (Goodwin, 1994; Seidel & Stürmer, 2014; Sherin & van Es, 2009). Thus, the question can be asked to what extent different constructs are transferable. Another question concerns the three components of diagnostic competences. It is important to know how these three components are interrelated to identify factors that teacher education can address in order to improve diagnostic competences of future teachers. To date, research approaches that consider the components altogether have hardly been pursued (Ohle & McElvany, 2015; see also Table 3). That further leads to the question of suitable instructional interventions that can be used to promote components of diagnostic competences. Teachers' professional knowledge is a crucial component, which has been emphasized many times in expertise research (e.g., Baumert et al., 2010; Fischer et al., 2012; Förtsch et al., 2016). Instructional support such as text-based interventions or lectures should consequently be effective with respect to knowledge development. Therefore, this dissertation investigates these instructional approaches (text-based intervention, lectures) to address debates about the effects of integrated instruction on the knowledge facets. By now, integrated instruction has been studied almost exclusively in the context of computer-based intervention studies (e.g., Evens et al., 2018; Harr et al., 2014, 2015; Janssen & Lazonder, 2016).

Therefore, the aims of this dissertation were:

1. … the validation of the video-based tool DiKoBi *Assess* that offered subject-specific challenges of biology instruction (i.e., content of the videos), including a whole-class perspective and addressing several subject-specific features of instructional quality. The validation referred to the content of the videos and the tasks used. As a second part of the validation, the construct of *diagnostic activities*, important for the analysis of components of diagnostic competences, was compared with the construct of *professional vision*.
   The first aim was addressed within Publication I and Publication II.
2. … the investigation of the relationships between the components professional knowledge, diagnostic activities, and diagnostic accuracy that constitute diagnostic competences according to the COSIMA framework.
   The second aim was addressed within Publication III.
3. … the analysis of the effects of different types of instructional support on the components of diagnostic competences. Of particular interest is the investigation of the effects of instructional support provided by a lecturer and via texts on the professional knowledge base.
   The third aim was addressed within Manuscript I and Manuscript II.

In Figure 4, an overview of the main aims is given.

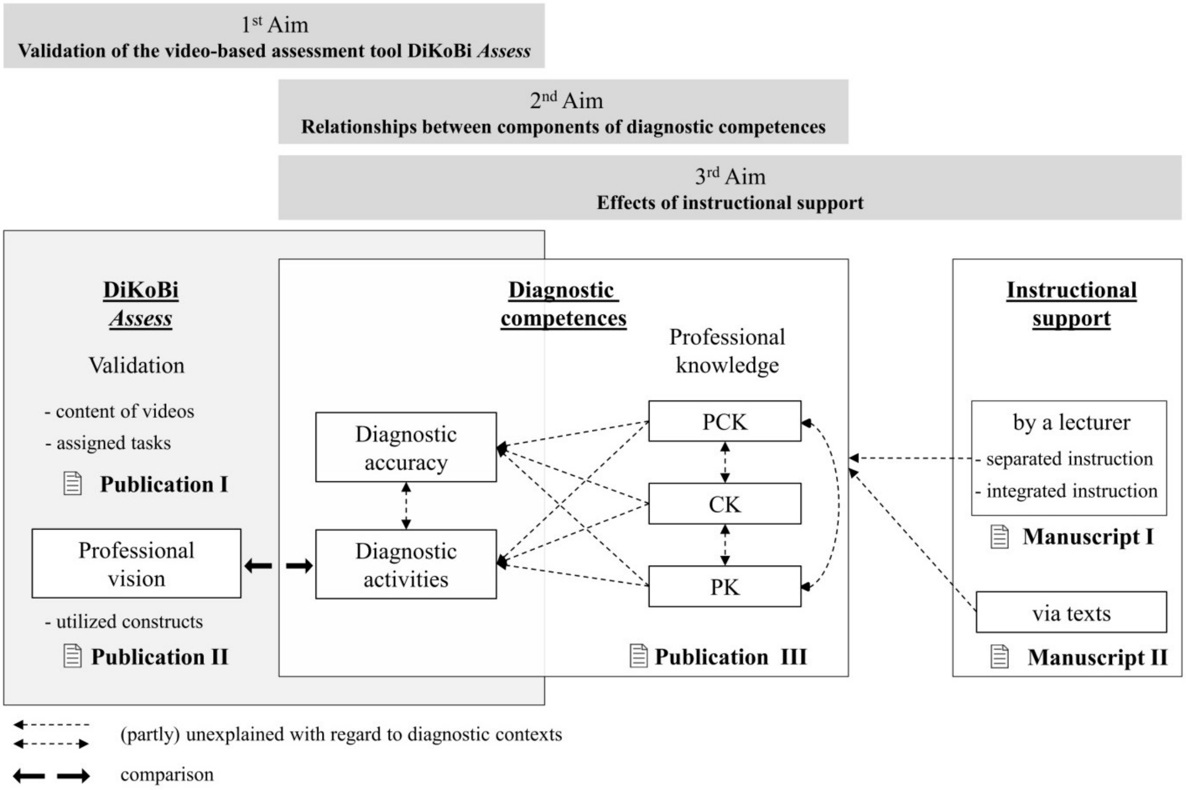

Figure 4. Overview of the main aims of this dissertation.

All three aims were addressed within the DFG-funded project COSIMA by developing, conducting, and analyzing three studies. Data for the validation process (cf. 1st aim), including the comparison of the constructs *diagnostic activities* and *professional vision,* was collected through a validation study with five in-service biology teachers (see Publication I and Publication II). Data for addressing the 2nd and the 3rd aim was collected with two main studies (see Figure 5). Study 1 was conducted with $N = 90$ pre-service biology teachers, Study 2 with $N = 103$ pre-service biology teachers. Both main studies were embedded within regular courses at the biology education of the LMU Munich. Data from the two studies were then considered in three separate analyses: For a cross-sectional analysis, data from both the pre-test of Study 1 and the pre-test of Study 2 were used (see Publication III). Subsequently, both pre- and post-test data from Study 1 and from Study 2 were analyzed separately. Manuscript I refers to the analysis of Study 2. Manuscript II refers to the analysis of Study 1. Within each publication/manuscript, an appropriate subsample was used depending on the goal of the analysis. Therefore, the sample number may differ from the total number of data used in the individual publications. Figure 5 gives an overview of the design of the studies.

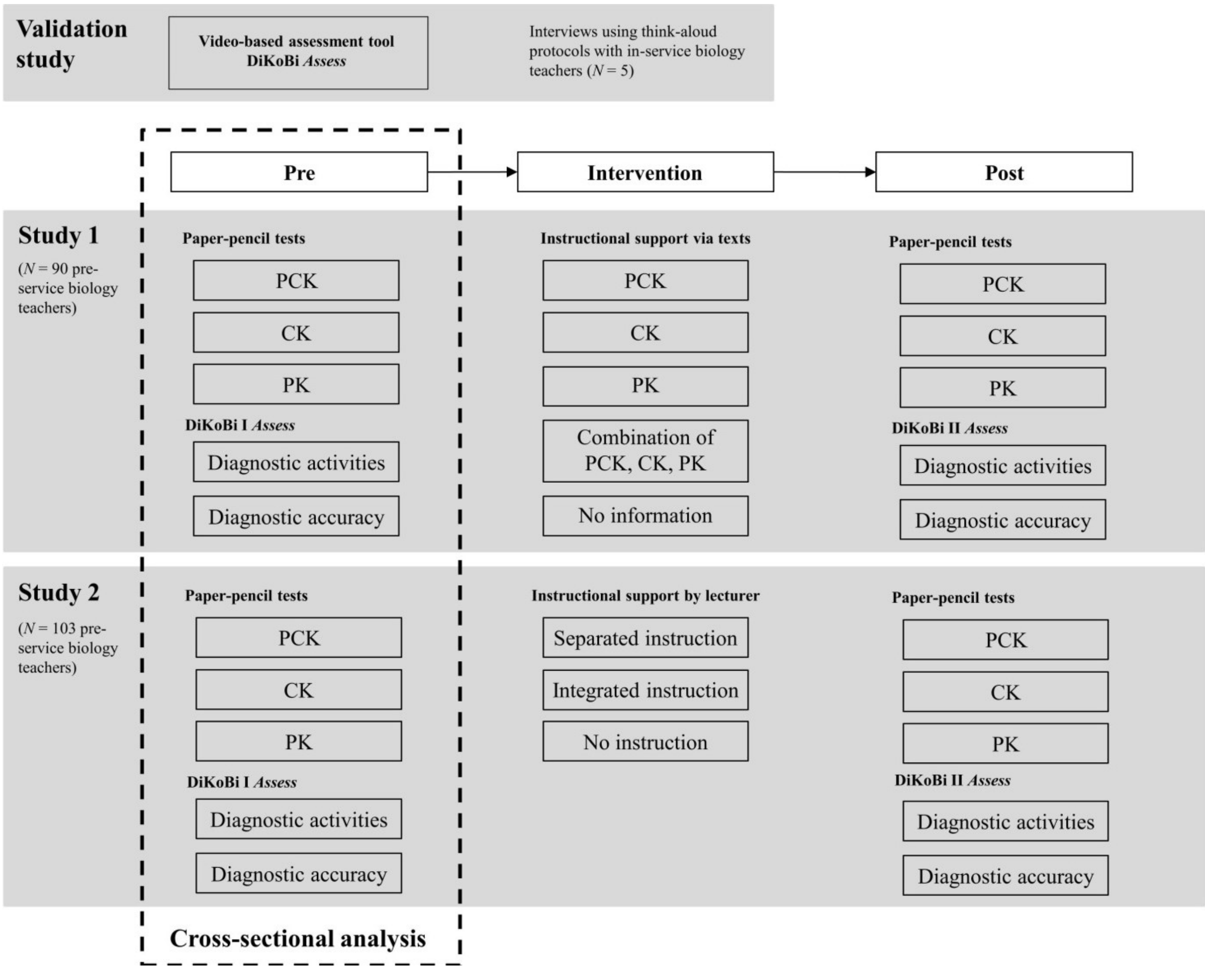

Figure 5. Design of the studies. In addition to Studies 1 and 2, a cross-sectional analysis was conducted with the pre-data from both main studies.

In both main studies, components of diagnostic competences were measured in a pre-post design. Professional knowledge was assessed with paper-pencil tests, diagnostic activities and diagnostic accuracy were assessed with the versions DiKoBi I *Assess* and DiKoBi II *Assess* of the video-based tool. Paper-pencil tests for measuring PCK and CK were adapted versions of the professional knowledge tests utilized in the ProwiN project, which proved to be valid (Jüttner et al., 2013; Jüttner & Neuhaus, 2013). The PCK and CK tests were adapted to the topic of *skin* since this was the topic that was also covered in the video-based assessment tool DiKoBi *Assess*. The paper-pencil test for measuring pre-service teachers' PK originated from the BilWiss project and was an adapted version of the short scale of the dimension *instruction* (Kunina-Habenicht et al., 2020; Kunter et al., 2014). This dimension covered generic characteristics of instructional quality with regard to the three basic dimensions *classroom management*, *supportive climate*, and *cognitive activation*, and was thus consistent with the theoretical considerations of the PK construct underlying the main studies (see Sections 1.1 and 1.3.1.1).

For the analysis of the professional knowledge tests, coding manuals were used by two independent raters. For analyzing diagnostic activities and diagnostic accuracy, a comprehensive coding manual was developed and used by three independent raters. As the coding manual was developed based on current literature, and all raters were trained and took part in regular discussions on complex coding cases, a valid measurement of diagnostic activities and diagnostic accuracy can be concluded (Arffman, 2016). Final results of two-way random intra-class correlations ($ICC_{absolute}$) showed a high agreement between the raters for both the professional knowledge tests and diagnostic activities/diagnostic accuracy (cf. Wirtz & Caspar, 2002). Additionally, fit indices of all measurement instruments indicated a reliable and objective measurement (cf. Bond & Fox, 2007; Boone et al., 2014; Boone & Staver, 2020; Wirtz & Caspar, 2002). Intra-class correlations and fit indices are reported in Publication III, Manuscript I, and Manuscript II.

3. Results

3.1 Publication I

Maria Kramer, Christian Förtsch, Julia Stürmer, Sonja Förtsch, Tina Seidel,
and Birgit J. Neuhaus

Measuring biology teachers' professional vision:
Development and validation of a video-based assessment tool

accepted for publication in
Cogent Education

Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, *7*(1).
https://doi.org/10.1080/2331186X.2020.1823155

3.2 Publication II

Maria Kramer, Christian Förtsch, Tina Seidel, and Birgit J. Neuhaus

Comparing two constructs for describing and analyzing teachers' diagnostic processes

accepted for publication in
Studies in Educational Evaluation

Kramer, M., Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation*, *28*. https://doi.org/10.1016/j.stueduc.2020.100973

3.3 Publication III

Maria Kramer, Christian Förtsch, William J. Boone, Tina Seidel, and Birgit J. Neuhaus

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

accepted for publication in
Education Sciences

Kramer, M., Förtsch, C., Boone, W. J., Seidel, T., & Neuhaus, B. J. (2021). Investigating Pre-Service Biology Teachers' Diagnostic Competences: Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy. *Education Sciences, 11*(3), 89. https://doi.org/10.3390/educsci11030089

3.4 Manuscript I

Maria Kramer, Christian Förtsch, and Birgit J. Neuhaus

Integrating or not-integrating – that is the question. Effects of integrated instruction on the development of pre-service biology teachers' professional knowledge

submitted for publication in
Frontiers in Education

Abstract

For successful classroom instruction, teachers require a well-founded knowledge base consisting of the three knowledge facets pedagogical-psychological knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK). However, there is not yet clarity about the circumstances and instructional pathways through which teachers can best develop these knowledge facets. In an experimental study ($N = 188$ pre-service biology teachers), we investigated the effects of separated instruction (knowledge facets were treated successively without linking) or integrated instruction (knowledge facets were presented in an interrelated way) in comparison to a control group that received no instruction for knowledge acquisition. Both pathways of instruction were provided by a lecturer on the curricular topic of senses and sensory organs, exemplified for the topic skin. Results point to the effectiveness of both ways of instruction in terms of knowledge increase for CK and PCK. However, descriptive results indicated that separated instruction was more effective for CK, but integrated instruction was more effective for PCK. No effects for PK could be found, possibly due to a ceiling effect. Overall, our results showed that instruction of CK within separate university courses can still be considered fruitful, while in terms of PCK more attention should be paid to integration.

Keywords: teacher education, science education, pedagogical content knowledge, instructional design, teacher professional knowledge.

Integrating or not-integrating – that is the question. Effects of integrated instruction on the development of pre-service biology teachers' professional knowledge

1 Introduction

Teacher education programs support pre-service teachers in acquiring professional knowledge that is fundamental for high-quality instruction (Baumert et al., 2010; Keller et al., 2017). Direct instructional guidance is one way to support knowledge acquisition and is therefore an important element of lectures or seminars that pre-service teachers attend at university. Direct instructional guidance can be understood as „providing information that fully explains the concepts and procedures that students are required to learn as well as learning strategy support that is compatible with human cognitive architecture" (Kirschner et al., 2006, p. 75). Concepts and procedures that pre-service teachers have to know about for powerful teaching mainly relate to three different facets of professional knowledge. These facets cover knowledge of subject-specific core ideas (content knowledge or CK), knowledge of subject-specific strategies to make subject-specific core ideas and content accessible for students (pedagogical content knowledge or PCK), and knowledge of general pedagogical-psychological principles and methods (pedagogical-psychological knowledge or PK) (Baumert & Kunter, 2013a; Shulman, 1987; Voss et al., 2011). Although there is evidence that linking and cross-referencing between knowledge facets is crucial for their retrievability and applicability, the knowledge facets are mainly addressed within separate university courses that rarely connect content and pedagogy (Harr et al., 2015; König et al., 2018; Renkl et al., 1996; Tröbst et al., 2019). For example, pre-service biology teacher attend courses in pedagogy, in which they are instructed in general teaching methods and strategies for classroom management (Voss et al., 2011). In courses of the discipline biology, pre-service teachers then acquire knowledge about specific biological topics (e.g., about the skin and its structure); whereas within didactical courses they get to know, for example, core ideas such as *structure and function*, and how to implement them in biology instruction (National Research Council, 2012; Sekretariat der Ständigen Konferenz der Kultusminister der Länderin der Bundesrepublik Deutschland [KMK], 2005). Pre-service teachers also get to know strategies for how to deal with student ideas and how to plan concept-oriented lessons that can all be considered as part of teachers' PCK. The integration of information from the three knowledge facets is largely the pre-service teachers' own task. In other words, how to deal with student ideas about the skin and its structure, and how to use scientific core ideas to foster students understanding of the specific content is rarely explicitly addressed. Since this task creates considerable difficulties and is hardly achieved successfully, researchers call for an integration of knowledge facets within teacher education, which makes the transformation into effective classroom instruction more likely (Ball, 2000; Kleickmann et al., 2017; Renkl et al., 1996; Tröbst et al., 2019). Studies focusing on the effects of integrated instruction in university courses, which consider all three knowledge facets for instructional input and as outcome measure, have hardly been conducted so far. To decide when and under which circumstances integration is appropriate, more empirical data from different domains is needed.

Therefore, the present study experimentally compared the effects of instructing all three knowledge facets, PK, CK, and PCK, in a separated or integrated way. The separated condition treats aspects of PK, CK, and PCK successively without linking and cross-referencing content, as it is usually done in university teacher education (Harr et al., 2014; König et al., 2018). In the integrated condition, a lesson planning model was used to structure instruction. For each phase of the planning model, corresponding aspects of all three knowledge facets were then presented in an interrelated way. Both conditions received direct instructional guidance of the knowledge facets through a lecturer. Thus, the chosen instructional approach allows at the same time a very practical investigation of the effects of integrated instruction since teacher training programs often include courses guided by a lecturer (Tröbst et al., 2019). Therefore, especially for designing teacher education programs, the present study is of great practical use as well.

1.1 Professional knowledge as part of teachers' professional competence

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

The importance of teachers' professional knowledge for instructional quality and student outcomes is empirically well-proven (Baumert et al., 2010; Fischer et al., 2012; Förtsch et al., 2016; Kunter et al., 2013). With regard to the subject and specific contextual and situational demands, different facets of professional knowledge have been distinguished on the basis of Shulman (1987). The common ground is that teachers need pedagogical-psychological knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK) during instruction (Baumert et al., 2010; Shulman, 1986, 1987; Voss & Kunter, 2013). Pedagogical-psychological knowledge (PK) is considered generic and independent from a specific subject, and has been conceptualized slightly differently. Whereas Voss et al. (2011) included knowledge about classroom processes and students' heterogeneity in their conceptualization of PK, König et al. (2014) considered knowledge about generic theories and methods of instruction and learning as well as of classroom management as essential for their conceptualization. In addition, classroom analyses offer a further perspective on PK. Research on instructional quality has identified three basic dimensions of instructional quality that occur more or less across domains: *classroom management*, *supportive climate*, and *cognitive activation*. *Classroom management* refers both to the structure and organization of instruction and to the management of students' behavior (Klieme et al., 2001; Praetorius et al., 2018; Schlesinger & Jentsch, 2016). *Supportive climate* (also often referred to as learning support) refers to the creation of a positive learning atmosphere in the classroom. It is characterized by a caring attitude of the teacher, a positive teacher-student relationship and other forms of support such as constructive feedback (Klieme et al., 2001; Lipowsky et al., 2009; Praetorius et al., 2018). *Cognitive activation* requires instruction that builds on students' prior knowledge and ideas, that uses challenging problems and questions to stimulate cognitive conflicts, that foster students' engagement in higher-level thinking and thus their in-depth elaboration of content, as well as students' active participation in classroom discourse (Baumert et al., 2010; Förtsch et al., 2017; Lipowsky et al., 2009). However, since many of the characteristics of cognitive activation have to be applied within subject-specific contexts, the operationalization of this basic dimension differs largely between studies due to content-specific issues of the study subject (Schlesinger & Jentsch, 2016). Furthermore, research findings revealed that "*classroom management* and *supportive climate* could be interpreted as basic conditions, which have to be established before implementing cognitively activating strategies of instruction" (Dorfner et al., 2018, p. 49). Thus, *cognitive activation* is more related to knowledge of subject matter (CK and PCK) (Baumert & Kunter, 2013b). In conclusion, knowledge about the basic dimensions *classroom management* and *supportive climate* as well as about corresponding strategies on how do deal with these dimensions, are important to create learning opportunities and an effective learning atmosphere in which pedagogical strategies and methods can be applied and adapted to student heterogeneity (cf. König & Kramer, 2016; Kunter et al., 2007). Applying PK in instruction can therefore also be seen as implementing the basic dimensions in instruction.

Furthermore, when teaching a subject, teachers need knowledge of subject matter. Professional knowledge referring to the understanding of subject-specific methods and core concepts is called content knowledge (CK) (Ball et al., 2008; Gess-Newsome, 2015). The knowledge that is necessary to make this content available for a particular group of students is referred to as pedagogical content knowledge (PCK), including aspects of content and pedagogy (Gess-Newsome, 2015; Shulman, 1987). Although conceptualizations of PCK differ, they all emphasized knowledge about subject-specific instruction and a student-related perspective as essential. Therefore, important components of PCK are knowledge about students' (mis)conceptions, as well as knowledge about subject-specific structures of instruction and corresponding teaching strategies (Depaepe et al., 2013; Park & Oliver, 2008; Schmelzing et al., 2013). With regard to subject-specific instructional quality (assessed through instructional quality features such as *cognitive activation*) and student achievement, researchers assumed teachers' subject-specific knowledge facets PCK and CK to be relevant (e.g., Kunter et al., 2013). While no direct effects of CK on *cognitive activation* have been found, there is evidence of the connectedness between CK and PCK (e.g., Krauss et al., 2008; Liepertz & Borowski, 2019). In contrast, PCK was shown to be highly predictive for instructional quality and students' achievement (Depaepe et al., 2013; Kulgemeyer et al., 2020; Schmelzing et al., 2013). An indirect effect of teachers' PCK on students' achievement mediated by *cognitive activation* was found, for example, in biology and mathematics education (Förtsch et al., 2016; Kunter et al., 2013). While there are conceptualizations in which both CK and PCK are included as subject-specific knowledge for teaching (e.g., Ball et al., 2008; Hill et al., 2008; Kersting et al., 2010), other researchers developed instruments for a separate measurement of the knowledge facets (e.g., Jüttner et al., 2013). Results from the COACTIV project (*Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers*) led to the conclusion that CK and PCK exist as two overlapping but distinct facets (Krauss et al., 2013). However, within the ProwiN project (*Professional Knowledge of Teachers in Science*) CK and PCK did not correlate significantly, and thus, could be measured independently (Förtsch et al., 2016).

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

Although we are aware of the advantage and necessity of considering both perspectives, the present study focuses more on the cognitive perspective on teacher professional knowledge that enables us to investigate professional knowledge quantitatively, thus, increasing the generalizability of the results.

1.2 Development of professional knowledge in teacher education

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

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

While the previous studies examined only two of the three knowledge facets, Evens et al. (2018) included all three knowledge facets in their study. Situated within the subject of French as a foreign language, one question they investigated was whether a learning environment in which PCK, PK, and CK are integrated is more effective for PCK development than a learning environment in which PCK, PK and CK are segregated. Five conditions (four experimental groups and one control group) differed in the knowledge facets that were presented, and in the way the knowledge facets were integrated. In contrast to Harr et al. (2014, 2015), Evens et al. (2018) found no significant differences between integrated and separated instruction. In both conditions, PCK increased moderately. However, whether both groups can equally apply their knowledge to the processing of practical examples was not investigated. Furthermore, results of the study support previous findings on the importance of explicit instruction on PCK for the development of subject-specific knowledge. The authors also emphasized that the instruction of all three knowledge facets might then be expedient "if teacher education aims at promoting not only teachers' PCK, but also their PK and CK" (Evens et al., 2018, p. 253). Therefore, addressing different knowledge facets and connecting them can still be considered appropriate to develop pre-service teachers' professional knowledge holistically.

1.3 The present study

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

Therefore, the main research question of the present study is: *Are there differences in the effectiveness of a separated or an integrated instruction on the development of pre-service teachers' professional knowledge facets (PK, CK, and PCK)*?

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

---

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

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

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

2 Materials and Methods

To completely reflect the methodology of the present study, the design of the original study is presented below. In the original study, we used two kinds of measuring instruments: paper-pencil tests to measure the knowledge facets and a video-based assessment tool to measure components of diagnostic competences. Since the video-based tool presented real-life classroom situations, we took the classroom context into account that is considered to play an important role when teachers have to apply their knowledge (Evens et al., 2018). Thus, measuring professional knowledge was based on the cognitive perspective, whereas measuring diagnostic competences was based on the situated perspective (cf. Hoth et al., 2016). However, since the video-based assessment tool was not part of this sub-analysis, further information on the video-based assessment tool is reported in Kramer et al. (2020).

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

2.1 Design and sample

The study had an experimental design and was embedded in a regular seminar held once a week. The seminar is attended by pre-service biology teachers at the beginning of their teacher education. In the seminar, pre-service teachers acquire knowledge about subject-specific theories and concepts for biology instruction. The study was conducted over two weeks in May 2019, with pre-testing and post-testing during the regular seminar time. The intervention was shifted to the weekend in between. Both pre- and post-tests took 50 minutes and included three paper-pencil tests each (see Figure 1). Thus, each knowledge facet was measured with a separate paper-pencil test, which was the same in pre- and post-test. Additionally, a video-based assessment tool was used to measure pre-service teachers' diagnostic competences in the pre- and post-test. However, measurements of the assessment tool were not considered for the present sub-analysis. For the intervention, pre-service teachers were randomly assigned to three different treatments. In treatment 1 and 2, a professional lecturer (first author of the article) gave three lectures on declarative and action-related aspects of PK, CK, and PCK relevant for the biological topic "skin". Each lecture took 90 minutes. During instruction, the different knowledge facets were addressed either in a separated or in an integrated way (see section *description of the treatments*). Pre-service teachers in treatment 3 (control group) did not receive any instruction. They only completed the pre- and the post-test. Informed consent documents stating an anonymous and voluntary participation were signed by all participants. The ethics committee of the Faculty of Psychology and Education of the LMU Munich approved the study in advance.

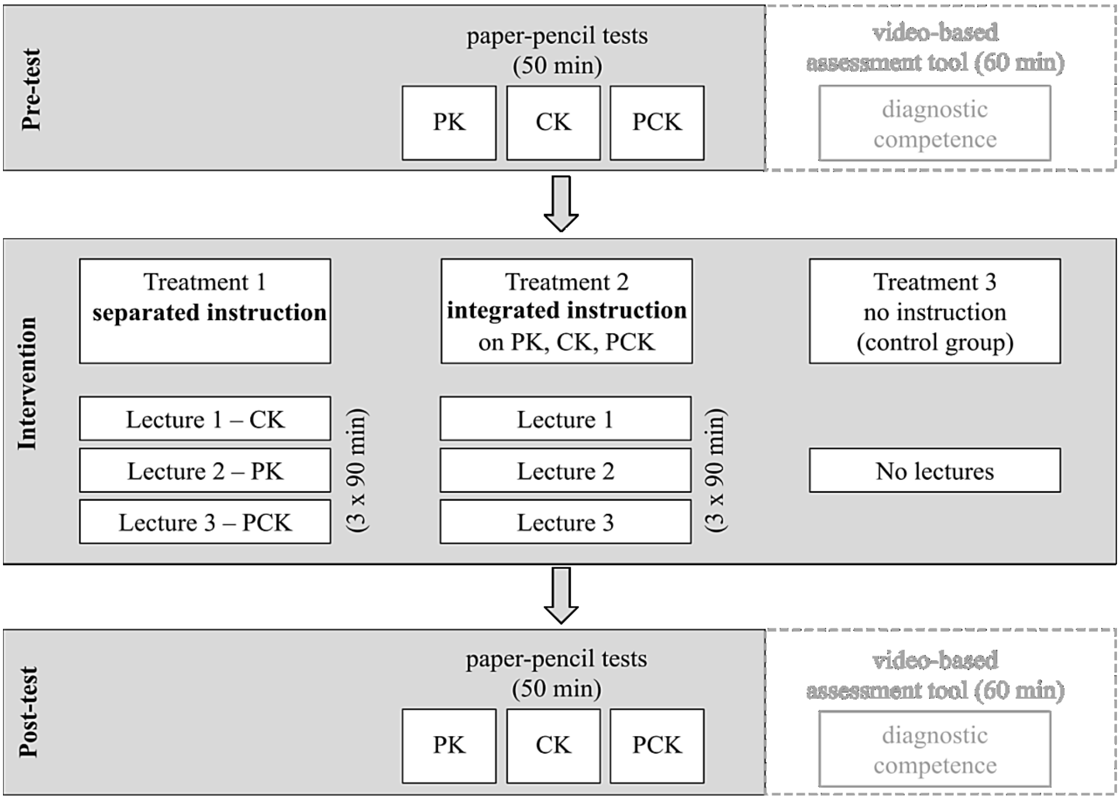

Figure 1. Design of the study. Measurements that were not considered in the present sub-analysis are framed with dashed lines.

The sample consisted of 118 pre-service biology teachers (66.9 % female; average study semester: $M = 3.02$, $SD = 1.20$; age in years: $M = 22.65$, $SD = 3.49$). 32.2 % of the pre-service teachers attended the academic track of teacher education, qualifying them for future teaching at German secondary schools ("*Gymnasium*"); 67.8 % attended programs for the non-academic track that qualifies students for vocational career. For an overview of the German school system, see Cortina and Thames (2013). Table 1 shows how the 118 participants were distributed among the three treatments. There was no statistically significant difference in age ($F(2,114) = 2.78$, $p = .07$), or percentage of pre-service teachers attending academic track ($F(2,115) = .53$, $p = .59$). They also did not statistically differ in their knowledge at the pre-test: PK_{pre} ($F(2,115) = .05$, $p = .96$), CK_{pre} ($F(2,115) = .03$, $p = .97$), PCK_{pre} ($F(2,115) = .87$, $p = .42$). However, a statistically significant difference was found in study semester ($F(2,115) = 3.94$, $p = .02$) between treatment 3 (control group) and treatment 2 (integrated instruction) as assessed by Tukey-HSD (mean difference control group-integrated instruction = .74, $p = .02$).

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

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

2.2 Description of the treatments

Pre-service teachers were randomly assigned to one out of three treatments: separated instruction (treatment 1), integrated instruction (treatment 2), or no instruction/control group (treatment 3). The intervention was held on a weekend, containing three 90-minutes lasting lectures for treatment 1 on the weekend's Saturday, and three 90-minutes lasting lectures for treatment 2 on Sunday. On both days of the weekend, the intervention was held by the same lecturer to reduce potential confounding effects. In addition, the lecturer, who was the first author of the article, prepared scripts to ensure that the overall content was kept constant in both treatments. Scripts were based on a review of the relevant literature and state-of-the-art research results of each knowledge facet with respect to the specific topic "skin". After intensive training, the lecturer strictly held the lecture according to the scripts to ensure that only the planned contents were addressed in the lectures. The distribution of the contents to the different lectures of the treatments is shown in Table 2.

Table 2. Overview of the content of the three lectures of each treatment

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

scientific inquiry methods, use of models, technical terms, linking…) - subject-specific theories (situational interest, conceptual change, …)	- phase 5: closing the lesson

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

In treatment 1, each of the lectures focused on one of the knowledge facets separately, meaning that the first lecture dealt with content knowledge (CK), the following lecture with pedagogical-psychological knowledge (PK), and the last lecture with pedagogical content knowledge (PCK). There was a 20-minute break between each lecture. In each of the three lectures of treatment 2, the knowledge facets were addressed in an integrated way. The overall structure of the three lectures of treatment 2 followed the basic structure of a planning model for biology instruction (see Dorfner et al., 2019), including different phases during the course of a lesson: introduction (with reactivation of students' prior knowledge, use of a focus question, and generating hypothesis), stage of elaboration and evaluation, backing up results, referring back and answering the focus question, as well as a consolidation of the content/concepts being taught. For each phase, corresponding aspects of CK, PK, and resulting PCK were presented. These aspects were the same for both treatments. Only the way of integrating the contents varied. Therefore, the planning model mentioned above was also considered in Treatment 1 as part of the lecture on PCK.

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

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

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

2.3 Professional knowledge tests

PK, CK, and PCK were measured through use of paper-pencil tests. In the tests, three different types of items have been utilized (open-ended, single best answer, multiple true/false). Open-ended items required a written response in a text field; single best answer (SBA) items required the selection of a single answer from a set of possible responses consisting of multiple distractors and one correct response; multiple true/false items required the assessment of each of four given responses with respect to their correctness (Campbell, 2011). Sample items are displayed in Table 4.

Table 4. Item examples of the CK and PCK test. Due to the different design of SBA-items, two examples are given.

Item type	**Item example**	**Corresponding knowledge facet**
Open-ended	Please evaluate the three-dimensional model by comparing as many different advantages and disadvantages as possible.	PCK
Single best answer (SBA)	*Read each statement and tick the statement if correct.* The dermis has subcutaneous fatty tissue and thus serves as protection against cold. ☐ During cornification, the cell interior dies and is replaced by the protein carotene. ☐ …	CK
	Please evaluate the statements below regarding the given teacher experiment. This experiment allows effective use of time. 1 = very ☐ 2 = rather ☐ 3 = neither ☐ 4 = rather not ☐ 5 = not at all ☐	PCK

Note: Information on items utilized in the PK test can be found by Kunina-Habenicht et al. (2020). 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 https://www.iqb.hu-berlin.de/fdz/studies/BilWiss.

PK was assessed by an adapted short version of a paper-pencil test utilized in the BilWiss project covering declarative knowledge of the dimension *instruction* (Kunina-Habenicht et al., 2020). According to Evens et al. (2018), PK includes at least knowledge about teaching methods and classroom management, which both were covered in the dimension *instruction* of the BilWiss project. The PK-test referred to basic dimensions of instructional quality containing items about generic features such as *classroom management*, *supportive climate*, and *cognitive activation* (Baumert et al., 2010; Klieme et al., 2001; Lipowsky et al., 2009), as well as items on general pedagogical issues of teaching such as teaching methods. The PK-measure included five SBA-items and ten multiple true/false items. Item scoring followed the instructions from the BilWiss project (Kunina-Habenicht et al., 2020). SBA-items were scored with either 0 points (wrongly ticked) or 2 points (correctly ticked). Multiple true/false items were scored with either 0 points (for 0 or 1 correctly ticked answers), 1 point (for 2 or 3 correctly ticked answers), or 2 points (for 4 correctly ticked answers, thus a completely solved task).

CK and PCK were assessed by adapted versions of the professional knowledge tests used in the ProwiN project (Jüttner et al., 2013; Jüttner & Neuhaus, 2013). The CK- and PCK-test covered declarative and action-related knowledge about the topic "skin" (in accordance with the knowledge that was addressed in the intervention and also covered in the video-based assessment tool). Based on the model of Tepner et al. (2012), the PCK-test covered two important components of biology teachers' PCK: knowledge of instructional strategies (model use and use of experiments) and knowledge of students' errors. The PCK-measure included eight open-ended items and five SBA-items. Open-ended items were scored in accordance with a coding manual that was adapted from Jüttner et al. (2013) and Jüttner and Neuhaus (2012), and written on the basis of the literature in science education. A maximum of 3 points could be achieved for each open-ended item. SBA-items represented ranking items, where statements to a given experiment were to be evaluated on a five-point Likert scale by the pre-service teachers (see Table 4). Prior to the scoring process, the items were rated by 16 in-service teachers who we considered as experts in biology education. In accordance with the tendency for correct answers the in-service biology teachers gave, we divided the Likert scale into positive, neither/nor, and negative parts for scoring pre-service teachers' ratings. For example, if the mean of the experts' rating was between 1 and 2, and a pre-service teacher check-marked 1 or 2, the answer was scored with 1 point. If the pre-service teacher check-marked 3, 4 or 5, the answer was scored with 0 points (cf. Jüttner et al., 2013). The CK-measure included 13 open-ended items and 15 SBA-items. Both open-ended items and SBA-items were scored in accordance with criteria provided in a coding manual adapted from Jüttner et al. (2013). A maximum of 3 points could be achieved for each open-ended item. SBA-items were scored with either 0 points (wrongly ticked) or 1 point (correctly ticked). To ensure objective and reliable coding, ten percent of both the PCK- and CK-test were coded by two independent raters utilizing the coding manuals. A high agreement between the two raters has been shown by the results of two-way random intra-class correlations (ICC_{absolute}): PCK: $ICC_{absolute}(310,310) = .84$, $p < .001$; CK: $ICC_{absolute}(341,341) = .97$, $p < .001$ (Wirtz & Caspar, 2002).

After item scoring, each knowledge test was analyzed separately using the Rasch partial credit model (PCM), which resulted in PK, CK, and PCK Rasch person measures (person abilities) for each pre-service teachers for each test instrument (Bond & Fox, 2007; Boone et al., 2014). The Rasch model also takes the difficulty of test items into account.

Person measures and item measures are calculated on the same scale using the unit "logits" as equal-interval units that allow comparisons between persons and items (Boone et al., 2014). For evaluating data fit, Outfit-MNSQ (mean-square) values, item reliability and person reliability for each test were used. According to Wright and Linacre (1994), item Outfit-MNSQ values below 1.5 indicate a productive measurement. Concerning the reliability values, high item reliability indicates that both the range of item difficulty and the sample size can be considered as appropriate to measure the items precisely. The person reliability describes the internal consistency of the measure. For example, a value of .50 means that the test discriminates the sample in 1 or 2 levels; higher values discriminate in more levels (Boone et al., 2014).

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

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

Knowledge facet	Number of items	outfit-MNSQ	person reliability	item reliability
PCK_{pre} (anchored)	13	< 1.44	.61	.93
PCK_{post}	13	< 1.15	.63	.95
CK_{pre} (anchored)	28	< 1.44	.69	.97
CK_{post}	28	< 1.43	.69	.97
PK_{pre} (anchored)	15	< 1.50	.57	.97
PK_{post}	15	< 1.19	.58	.98

2.4 Analyses

First, raw data of all knowledge tests (pre/post: PK, CK, PCK) was analyzed using the Rasch PCM (Bond & Fox, 2007) with the software Winsteps 3.81 (Linacre, 2014) to calculate person measures. The resulting equal-interval person measures were used for all following analyses. Second, descriptive results and Pearson's correlations were calculated utilizing IBM

SPSS Statistics (version 26) to describe the development and correlation between the knowledge facets. SPSS has also been utilized for running mixed ANOVAs separately for each knowledge facet to reveal possible time effects and interaction effects between time and treatments. Assumptions for applying mixed ANOVAs were checked. There was a violation of normal distribution for PCK_{pre} and PK_{post} of treatment 2 (integrated instruction), as well as for PK_{post} of the control group, as assessed by the Shapiro-Wilk test (p > .05). For PK and CK, homogeneity of the error variances, as assessed by Levene's test ($p > .05$), as well as homogeneity of covariances, as assessed by Box's test (PK: $p = .53$; CK: $p = .19$) were given. However, for PCK_{post}, Levene's test was significant ($p = .03$). We therefore focused on Tukey-HSD post-hoc comparisons and calculated repeated measures ANOVAs for each treatment.

As an effect size measure, we used partial η^2, applying the following benchmark values: .01 for small effects, .06 for medium effects, and .14 for large effects (Cohen, 1988; Richardson, 2011).

3 Results

3.1 Descriptive statistics and correlations

An overview of the descriptive results of all measurements can be found in Table 6. For each knowledge facet, the mean values increased between pre-test and post-test. However, for PCK and CK, the increases were greater. Additionally, it is worth mentioning that for PK the control group showed the largest increase compared to treatment 1 or 2. Pearson's correlations between the knowledge facets showed that PK, as well as CK and PCK, could mostly be measured independently. For the pre-test, there were small correlations between CK_{pre} and PCK_{pre} ($r = .29$, $p < .01$), as well as between PCK_{pre} and PK_{pre} ($r = .28$, $p < .01$). For the post-test, there were a medium correlation between CK_{post} and PCK_{post} ($r = .40$, $p < .01$), and a small correlation between PCK_{post} and PK_{post} ($r = .24$, $p < .01$).

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

	Treatment 1 separated instruction *N* = 42		**Treatment 2 integrated instruction** *N* = 40		**Treatment 3 control group** *N* = 36	
PK [a]	Pre	Post	Pre	Post	Pre	Post
M	5.27	5.36	5.28	5.33	5.24	5.38
SD	.49	.51	.56	.60	.52	.52
Min	4.21	4.34	4.38	4.34	4.02	4.66
Max	6.49	6.48	6.49	7.28	6.21	7.28
CK [a]						
M	4.49	5.18	4.44	4.98	4.46	4.53
SD	.68	.81	.78	.76	.81	.67
Min	3.30	3.44	2.26	3.19	1.74	3.19
Max	5.85	6.68	5.70	6.37	5.99	5.66
PCK [a]						
M	3.83	4.37	3.62	4.35	3.79	4.23
SD	.69	.55	.75	.62	.87	.81
Min	2.62	3.19	1.27	2.83	1.32	2.83
Max	5.28	5.53	4.65	5.36	5.28	5.70

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

3.2 Effects on PK, CK, and PCK

For PK, results of calculating the mixed ANOVA showed a significant effect of time ($F(1, 115) = 3.94$, $p = .05$, partial $\eta^2 = .03$), but no interaction effect between time and treatments ($F(2, 115) = .23$, $p = .79$, partial $\eta^2 < .01$), meaning that pre-service teacher acquired more PK regardless of treatment affiliation. However, due to the high person measures in both the pre- and the post-test, a ceiling effect can be stated (see Figure 2).

Mixed ANOVAs for CK revealed that there was a statistically significant effect of time for treatment 1 (separated instruction) ($F(1,41) = 28.64$, $p < .01$, partial $\eta^2 = .41$) and for treatment 2 (integrated instruction) ($F(1,39) = 33.44$, $p < .01$, partial $\eta^2 = .46$). There was no significant effect of time for the control group ($F(1,35) = .39$, $p = .54$, partial $\eta^2 = .01$). Furthermore, the interaction effect between time and treatment was significant in terms of

CK-development ($F(2,115) = 7.66$, $p < .01$, partial $\eta^2 = .12$). However, Tukey-HSD post-hoc tests of the mixed ANOVA revealed no significant difference between the groups. Despite missing significance, calculation of Tukey-HSD indicated that the separated instruction might have had an advantage over the control group (mean difference separated instruction-control group = .34, $p = .07$). In contrast, this potential advantage was not apparent for the integrated instruction (mean difference integrated instruction-control group = .22, $p = .33$). However, to understand the significant interaction effect, we run another analysis. Since pre-service teachers did not statistically differ in their CK_{pre}, we also analyzed group differences in the post-test and calculated a one-way ANOVA for CK_{post}. In the post-test, CK-scores differed significantly between the treatments. According to the Tukey-HSD post-hoc tests, control group differed significantly from treatment 1 (separated instruction) (mean difference control group-separated instruction = -.65, $p = .001$), and treatment 2 (integrated instruction) (mean difference control group-integrated instruction = -.44, $p = .03$) at the post-test. Overall, the results showed the effectiveness of the lectures, and might indicate a greater potential of the separated instruction in terms of CK development (see Figure 2).

Mixed ANOVAs for PCK could not be interpreted due to the significance of the Levene's test ($p = .03$), which still remained after cox-box-powertransformation (Hemmerich, 2016). Additionally, Tukey-HSD post-hoc tests showed no statistically significant difference between the treatments. However, repeated measures ANOVA for the total sample (without consideration of treatments) revealed a significant difference between PCK_{pre} and PCK_{post} ($F(1,115) = 77.76$, $p < .01$, partial $\eta^2 = .40$). Therefore, we run repeated measures ANOVA for each treatment separately. Results showed significant increases from PCK_{pre} to PCK_{post} for all three treatments (see Figure 2), which were much more pronounced for the treatments 1 and 2, which included instruction (treatment 1: $F(1,41) = 30.47$, $p < .01$, partial $\eta^2 = .43$; treatment 2: $F(1,39) = 44.72$, $p < .01$, partial $\eta^2 = .53$; control group: $F(1,35) = 11.38$, $p < .01$, partial $\eta^2 = .25$). Since the highest estimate of explained variance was reported for treatment 2, the integrated instruction might be more beneficial to foster knowledge acquisition in terms of PCK than addressing knowledge facets separately.

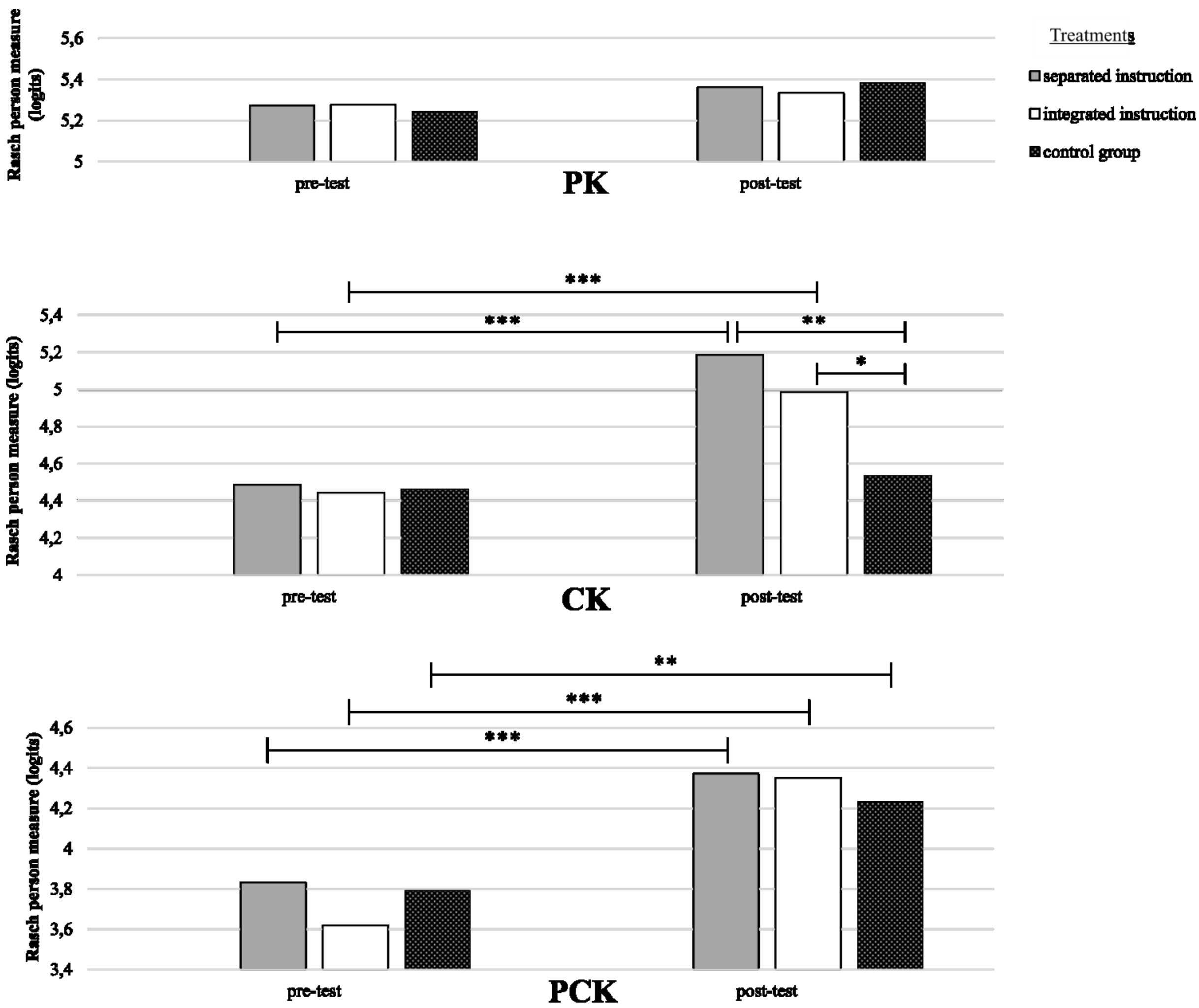

Figure 2. Comparison of pre- and post-test person measures between the different treatments, illustrated for each knowledge facet (***p < .001, **p < .01, *p < .05). For PK and CK, mixed ANOVAs have been performed, whereas we run repeated ANOVAs for PCK. Note that the person measures were calibrated for each knowledge facet separately, and therefore, they cannot be compared with each other.

4 Discussion

In the present study, we compared the effects of separated or integrated instruction on PK, CK, and PCK on the development of pre-service teachers' professional knowledge. Of critical significance was that all three knowledge facets were addressed both as part of the instruction and as outcome variables, whereas previous studies were often limited to specific knowledge facets (cf. Janssen & Lazonder, 2016; Harr et al., 2014). Regarding this point, our study provides insights into how teachers' professional knowledge can be fostered. Furthermore, we investigated the integration of the knowledge facets in a very practical way within regularly scheduled lectures. Therefore, the study also provides practical value for realizations of teacher education programs. Although the use of computer-based learning environments offers advantages in terms of controllability and standardizability, we decided to investigate the integration of the knowledge facets in lectures in order to expand methodological approaches within the investigation of knowledge instruction. In addition, study results could be used directly to adapt instruction due to the affiliation to our courses. Therefore, it addressed how curricular content should be presented in courses and lectures of science teacher education. Despite the evidence of greater effectiveness of integrated instruction (Harr et al., 2015; Janssen & Lazonder, 2016), knowledge facets are largely taught in separate courses in teacher education. Since curricular restructuring is not feasible without enormous effort, the purpose of this study was to generate more evidence on the extent to which restructuring in lectures might be effective. For this purpose, three treatments (separated instruction, integrated instruction, and control group receiving no instruction) were compared on participants' PK, CK, and PCK development.

The following main findings of the statistical analysis can be noted: All knowledge facets (PK, CK, PCK) increased from pre- to post-test. The largest increases were shown with respect to the development of CK and PCK. However, not all increases can be attributed to the intervention. Referring to PK, a small time effect but no interaction effect between time and treatments could be found. Instead, we noticed a ceiling effect, indicating that the utilized PK-test has been too easy and did not differentiate the sample enough (Linacre, 2014). In the future, we should therefore raise test difficulty through an extension of the PK-test by including other scales of the original BilWiss project (Kunina-Habenicht et al., 2020). Thus, at this stage, we cannot make any firm conclusions about differences in the effectiveness of separated or integrated instruction on the development of PK.

Referring to the development of CK, both addressing knowledge facets separately or integrated in lectures were effective. There were large interaction effects between time and treatments of both the separated and the integrated instruction compared to the control group that showed no significant difference between CK_{pre} and CK_{post}. Furthermore, the descriptive results indicated that the separated instruction was slightly more effective for CK development. This seems plausible because before teachers can teach a topic, they first need knowledge about the subject matter and they have to understand the underlying core concepts before actually planning instruction (Ball et al., 2008; Kleickmann et al., 2017). Furthermore, pre-service teachers can be considered as novices for whom less integration of knowledge facets is characteristic, in contrast to the knowledge of experienced in-service teachers, which is stronger integrated and encapsulated (Krauss et al., 2008). Therefore, due to little subject-specific knowledge about a specific topic and due to the still less integrated knowledge structures, the separated instruction might be the more effective one in terms of CK development. Here, future studies that take the different levels of prior knowledge of different pre-service teachers into account could provide more differentiated insights on the circumstances of a beneficial instruction on CK.

Referring to the development of PCK, it was striking that a large time effect was found not only in the two treatment groups but also in the control group who did not receive any kind of instruction during the intervention. This effect may be due to the processing of the biology-specific classroom situations provided in the video-based assessment tool that we used to measure diagnostic competences in the pre- and post-test within the original study. Even though these measurement results are not part of the present sub-analysis, the use of the tool seems to have had an impact on the knowledge facets captured as defined by the cognitive perspective. To explain this observation, we want to refer to the Refined Consensus Model mentioned in the theoretical section to define the PCK-constructs used in our study more precisely. Thus, the paper-pencil tests utilized in the pre- and post-tests measured pre-service teachers' individual subject-specific knowledge (*personal* PCK) that reflected their person-specific reservoir of declarative and action-related knowledge as well as the teachers' own teaching and learning experiences (Carlson et al., 2019). When elaborating on the PCK aspects of the intervention, we focused on topic-specific literature within the field and state-of-the-art research results. Thus, this form of PCK represented *collective* PCK that interacted with pre-service teachers' *personal* PCK and was therefore assumed to change it. Additionally, we assume that pre-service teachers had to rely on their *personal* PCK when working in the video-based assessment tool during the diagnostic process while simultaneously drawing on *enacted* PCK that is generated in the moment of action (Alonzo et al., 2019). Even if pre-service teachers were not actually in action themselves, they engaged in the practice of science teaching in terms of reflecting on biology instruction, and thus, utilized *enacted* PCK. It is assumed that through reflection, *enacted* PCK can be transformed to *personal* PCK, and thus, the experiences can become part of future knowledge (Alonzo et al., 2019). Since we used videos of real-life classroom instruction, the videos and their reflection might have elicited knowledge which could then be accessed in the post-test. This could have been, for example, knowledge about the use of three-dimensional models triggered by the situated context in the video. The videos thus functioned as an additional prompt for the retrieval of *personal* PCK (Alonzo & Kim, 2016; Kersting et al., 2010; Seidel & Stürmer, 2014). Therefore, the results of our sub-analysis also emphasized the relevance of considering both, a cognitive and a situated perspective on teacher professional knowledge. Considering the results, there is the possibility that an interaction effect between time and treatments was overshadowed by the effect of the video-based assessment tool. However, effect sizes suggest higher effects for the two treatments that received instruction, with the greatest value of explained variance for the integrated instruction. Therefore, our findings indicate a benefit for the development of PCK when knowledge facets are addressed in an interrelated way, which is in line with previous research findings (Harr et al., 2014, 2015). Through an integrated instruction on the knowledge facets, PCK development might benefit from the interaction of PK and CK, as well as from the explicit instruction on PCK itself (Kleickmann et al., 2017; Krauss et al., 2008; Schneider & Plasman, 2011; Tröbst et al., 2019).

Despite these insights, there are some limitations to the present study. First, some assumptions relevant for applying mixed ANOVAs were violated. However, the violation of normal distribution for PCK_{pre} and PK_{post} of treatment 2 (integrated instruction), as well as for PK_{post} of the control group, can be explained by the fact that PCK_{pre} was rather difficult for the students. Thus, due to the small group sizes of treatments, the probability that the normal distribution is violated increased. Results of PCK_{post} tests were normal distributed for all treatments. PK_{post}, on the other hand, was too easy, which is why there was no normal distribution for two treatments there either. A second limitation concerns the knowledge tests used. Since we used the same paper-pencil test for pre- and post-test, a test effect might have played a role concerning test results. Additionally, knowledge increases of the knowledge facets, even when not significant, might be due to pre-service teachers' regular courses that took place during the study but were not part of it. Regarding the ceiling effect resulting from the utilized PK-test, we cannot make any statements on the effects of the treatments on PK, which is why the discussion refers primarily to CK and PCK findings. Therefore, an extended test version should be piloted and used in future studies. Accordingly, the current study design would have to be applied again in a similar way using an extended PK test version in order to conduct and discuss the targeted holistic investigation of all three knowledge facets. This is especially important to evaluate whether integrated instruction might also be favorable in terms of PK development. If this is the case, university courses would need to pursue a stronger interlocking of the more practice-oriented facets PK and PCK (cf. König et al., 2018). At the same time, expanding the study could result in a larger sample size of the individual treatments. The increase in sample size might then provide more clarity about the partly descriptive trends in PCK and CK. Conducting the study at other universities would also be conceivable. However, it is a great challenge to embed the lecture-based study in exactly the same form at other universities since conception and implementation were strongly oriented towards the instructional practice of the biology education at the LMU Munich. However, this also represents a strength, as research and practice have been combined.

In addition, conceptual limitations regarding the knowledge tests must also be considered. Thus, we could only measure specific aspects of a pre-service teacher's knowledge facet referring to articulable knowledge "related to the teaching and learning of specific science topics" (Alonzo et al., 2019, p. 273). The paper-pencil tests did not allow us to capture more dynamic forms of knowledge that are used in practice, quasi "in action" (Alonzo & Kim, 2016), and thus, the measurement of professional knowledge with paper-pencil tests might lack sufficiency (Liepertz & Borowski, 2019). Measurement data from the video-based tool will provide more information here. Nonetheless, with reference to the RCM, for example, it remains open which particular form of PCK was addressed, and to what extent integrated instruction is equally effective for all forms. Consequently, future studies would do well not only to differentiate into knowledge facets, but also to consider potential subforms within a knowledge facet (as in the case of the PCK forms) that is relevant to interpretation and eventual consequences.

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

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

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Conceptualization was done by MK, CF, and BJN. Formal analysis, and writing of the manuscript was done by MK. Critical editing of the manuscript was done by CF and BJN.

Funding

The research for this article was funded by the German Research Association (Deutsche Forschungsgemeinschaft, DFG) (FOR2385, NE 1186/8-1).

5 References

Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the Complexity of Science Teachers' PCK in Action: Enacted and Personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (pp. 271–286). Springer.

Alonzo, A. C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, *53*(8), 1259–1286. https://doi.org/10.1002/tea.21271

Alonzo, A. C., & Kim, J. (2018). Affordances of video-based professional development for supporting physics teachers' judgments about evidence of student thinking. *Teaching and Teacher Education*, *76*, 283–297. https://doi.org/10.1016/j.tate.2017.12.008

Ball, D. L. (2000). Bridging Practices: Intertwining Content and Pedagogy in Teaching and Learning to Teach. *Journal of Teacher Education,*, *51*(3), 241–247. https://doi.org/10.1177/0022487100051003013

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Baumert, J., & Kunter, M. (2013a). The COACTIV Model of Teachers' Professional Competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–48). Springer.

Baumert, J., & Kunter, M. (2013b). The Effect of Content Knowledge and Pedagogical Content Knowledge on Instructional Quality and Student Achievement. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 175–205). Springer.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift Für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Bond, T. G., & Fox, C. M. F. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum Associates.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2013). The Diagnostic Skills of Mathematics Teachers. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the* *mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 229–248). Springer.

Campbell, D. E. (2011). How to write good multiple-choice questions. *Journal of Paediatrics and Child Health*, *47*(6), 322–325. https://doi.org/10.1111/j.1440-1754.2011.02115.x

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (pp. 77–94). Springer.

Cauet, E., Liepertz, S., Borowski, A., & Fischer, H. E. (2015). Does it Matter What We Measure? Domain-specific Professional Knowledge of Physics Teachers. *Swiss Journal of Educational Research*, *37*(3), 463–480. https://doi.org/10.24452/sjer.37.3.4963

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

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

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12–25. https://doi.org/10.1016/j.tate.2013.03.001

Dorfner, T., Förtsch, C., & Neuhaus, B. J. (2018). Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction. *Learning and Instruction*, *56*, 42–53. https://doi.org/10.1016/j.learninstruc.2018.03.001

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

Evens, M., Elen, J., Larmuseau, C., & Depaepe, F. (2018). Promoting the development of teacher professional knowledge: Integrating content and pedagogy in teacher education. *Teaching and Teacher Education*, *75*, 244–258. https://doi.org/10.1016/j.tate.2018.07.001

Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional Knowledge of Science Teachers. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second International*

Handbook of Science Education (Vol. 85, pp. 435–448). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9041-7_30

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M., Girwidz, R., Obersteiner, A., Reiss, K., Stürmer, K., Siebeck, M., Schmidmaier, R., Seidel, T., Ufer, S., Wecker, C., & Neuhaus, B. J. (2018). Systematizing Professional Knowledge of Medical Doctors and Teachers: Development of an Interdisciplinary Framework in the Context of Diagnostic Competences. *Education Sciences*, *8*(4), 207. https://doi.org/10.3390/educsci8040207

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

Förtsch, C., Werner, S., Kotzebue, L. von, & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, *38*(17), 2642–2666. https://doi.org/10.1080/09500693.2016.1257170

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

Harr, N., Eichler, A., & Renkl, A. (2014). Integrating pedagogical content knowledge and pedagogical/psychological knowledge in mathematics. *Frontiers in Psychology*, *5*, 1–10. https://doi.org/10.3389/fpsyg.2014.00924

Harr, N., Eichler, A., & Renkl, A. (2015). Integrated learning: ways of fostering the applicability of teachers' pedagogical and psychological knowledge. *Frontiers in Psychology*, *6*, 1–16. https://doi.org/10.3389/fpsyg.2015.00738

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., Schmidmaier, R., Neuhaus, B. J., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., & Fischer, F. (2019). Facilitating Diagnostic Competences in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education. *Frontline Learning Research*, *7*, 1–24. https://doi.org/10.14786/flr.v7i4.384

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

Hemmerich, W. A. (2016). *StatistikGuru: Box-Cox Powertransformation berechnen*. https://statistikguru.de/rechner/box-cox.html [Accessed December 17, 2020].

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students. *Journal for Research in Mathematics Education*, *39*(4), 372–400.

Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM*, *48*(1-2), 41–53. https://doi.org/10.1007/s11858-016-0759-y

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

Janssen, N., & Lazonder, A. W. (2016). Supporting pre-service teachers in designing technology-infused lesson plans. *Journal of Computer Assisted Learning*, *32*(5), 456–467. https://doi.org/10.1111/jcal.12146

Jüttner, M., Boone, W. J., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, *25*(1), 45–67. https://doi.org/10.1007/s11092-013-9157-y

Jüttner, M., & Neuhaus, B. J. (2012). Development of Items for a Pedagogical Content Knowledge Test Based on Empirical Analysis of Pupils' Errors. *International Journal of Science Education*, *34*(7), 1125–1143. https://doi.org/10.1080/09500693.2011.606511

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

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

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

Kind, V., & Chan, K. K. H. (2019). Resolving the amalgam: connecting pedagogical content knowledge, content knowledge and pedagogical knowledge. *International Journal of Science Education*, *41*(7), 964–978. https://doi.org/10.1080/09500693.2019.1584931

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

Kleickmann, T., Tröbst, S., Heinze, A., Bernholt, A., Rink, R., & Kunter, M. (2017). Teacher Knowledge Experiment: Conditions of the Development of Pedagogical Content Knowledge. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Methodology of Educational Measurement and Assessment. Competence assessment in education: Research, models and instruments* (Vol. 59, pp. 111–129). Springer International Publishing. https://doi.org/10.1007/978-3-319-50030-0_8

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I:

Aufgabenkultur und Unterrichtsgestaltung [Mathematics Instruction in Secondary Education: Task Culture and Instructional Processes]. In Bundesministerium für Bildung und Forschung (Ed.), *TIMMS - Impulse für Schule und Unterricht [TIMSS – impetus for school and teaching]* (pp. 43–57). Bundesministerium für Bildung und Forschung (BMBF).

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, *38*, 76–88. https://doi.org/10.1016/j.tate.2013.11.004

König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M., Strauß, S., Bremerich-Vos, A., Fladung, I., & Kaiser, G. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? [General pedagogical knowledge versus pedagogical content knowledge?]. *Zeitschrift für Erziehungswissenschaft, 21*(3), 1–38. https://doi.org/10.1007/s11618-017-0765-z

König, J., & Kramer, C. (2016). Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM*, *48*(1-2), 139–151. https://doi.org/10.1007/s11858-015-0705-4

König, J., Lammerding, S., Nold, G., Rohde, A., Strauß, S., & Tachtsoglou, S. (2016). Teachers' Professional Knowledge for Teaching English as a Foreign Language. *Journal of Teacher Education*, *67*(4), 320–337. https://doi.org/10.1177/0022487116644956

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Holt, Reinhardt & Winston.

Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, *7*(1). https://doi.org/10.1080/2331186X.2020.1823155

Krauss, S., Blum, W., Brunner, M., Neubrand, M., Baumert, J., Kunter, M., Besser, M., & Elsner, J. (2013). Mathematics Teachers' Domain-Specific Professional Knowledge: Conceptualization and Test Construction in COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 147–174). Springer.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716–725. https://doi.org/10.1037/0022-0663.100.3.716

Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., Riese, J., Schecker, H., Schröder, J., & Vogelsang, C. (2020). Professional knowledge affects action-related skills: The development of preservice physics teachers' explaining skills during a field experience. *Journal of Research in Science Teaching*, *4*(3), 1105. https://doi.org/10.1002/tea.21632

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

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

Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., Krauss, S., Löwen, K., Neubrand, M., & Tsai, Y.-M. (2007). Linking Aspects of Teacher Competence to their Instruction: Results from the COACTIV Project. In M. Prenzel (Ed.), *Studies on the educational quality of schools: The final report on the DFG Priority Programme* (pp. 32–52). Waxmann.

Liepertz, S., & Borowski, A. (2019). Testing the Consensus Model: relationships among physics teachers' professional knowledge, interconnectedness of content structure and student achievement. International Journal of Science Education, 41(7), 890–910. https://doi.org/10.1080/09500693.2018.1478165

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

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, *19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press. https://doi.org/10.17226/13165

Park, S., & Oliver, J. S. (2008). Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Research in Science Education*, *38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert Knowledge: Analyses and Remedies. *Educational Psychologist*, *31*(2), 115–121.

Richardson, J. T.E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM*, *48*(1-2), 29–40. https://doi.org/10.1007/s11858-016-0765-0

Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the 'cardiovascular system'. *International Journal of Science and Mathematics Education*, *11*(6), 1369–1390.

Schneider, R. M., & Plasman, K. (2011). Science Teacher Learning Progressions: A Review of Science Teachers' Pedagogical Content Knowledge Development. *Review of Educational Research*, *81*(4), 530–565.

Scott, P., Mortimer, E., & Ametller, J. (2011). Pedagogical link-making: a fundamental aspect of teaching and learning scientific conceptual knowledge. *Studies in Science Education*, *47*(1), 3–36. https://doi.org/10.1080/03057267.2011.549619

Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, *51*(4), 739–771. https://doi.org/10.3102/0002831214531321

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

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

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Havard Educational Review*, *57*, 1–22.

State Institute of School Quality and Educational Research, Munich. (2017). *LehrplanPLUS Mittelschule: Natur und Technik 5*. https://www.lehrplanplus.bayern.de/fachlehrplan/mittelschule/5/nt/regelklasse#77436 [Accessed December 21, 2020].

Tarman, B. (2016). Discipline or Classroom Management. *Journal of Learning and Teaching in Digital Age*, *1*(2), 37–44.

Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., Hubertina, T., & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften [Item Development Model for Assessing Professional Knowledge of Science Teachers]. *Zeitschrift Für Didaktik Der Naturwissenschaften*, *18*.

Tröbst, S., Kleickmann, T., Depaepe, F., Heinze, A., & Kunter, M. (2019). Effects of instruction on pedagogical content knowledge about fractions in sixth-grade mathematics on content knowledge and pedagogical knowledge. *Unterrichtswissenschaft*, *47*(1), 79–97. https://doi.org/10.1007/s42010-019-00041-y

Voss, T., & Kunter, M. (2013). Teachers' General Pedagogical/Psychological Knowledge. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 207–227). Springer. https://doi.org/10.1007/978-1-4614-5149-5_10

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, *103*(4), 952–969. https://doi.org/10.1037/a0025125

Wadouh, J., Liu, N., Sandmann, A., & Neuhaus, B. J. (2014). The Effect of Knowledge linking Levels in Biology Lessons upon Students' Knowledge Structure. *International Journal of Science and Mathematics Education*, *12*, 25–47. https://doi.org/10.1007/s10763-012-9390-8

Wirtz, M. A., & Caspar, F. (2002). *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.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(370). https://www.rasch.org/rmt/rmt83b.htm [Accessed June 06, 2020].

3.5 Manuscript II

Maria Kramer, Christian Förtsch, and Birgit J. Neuhaus

Can pre-service biology teachers' professional knowledge and diagnostic activities be fostered by self-directed knowledge acquisition via texts?

submitted for publication in
Education Sciences
(Special Issue "Scientific Reasoning in Science Education:
From Global Measures to Fine-Grained Descriptions of Students' Competencies")

education sciences

UNDER REVIEW

MDPI

Article

Can pre-service biology teachers' professional knowledge and diagnostic activities be fostered by self-directed knowledge acquisition via texts?

Maria Kramer[1,*], Christian Förtsch[1] and Birgit J. Neuhaus[1]

[1] Biology Education, Department I, Faculty of Biology, LMU Munich, Munich, Germany; christian.foertsch@bio.lmu.de (C.F.); didaktik.biologie@lrz.uni-muenchen.de (B.J.N.)
* Correspondence: maria.kramer@bio.lmu.de (M.K.)

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

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

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Educ. Sci.* **2021**, *11*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

1. Introduction

Scientific reasoning, as a component of scientific inquiry, encompasses reasoning and problem-solving processes that count as crucial for coping with science-related issues of everyday-life [1,2]. Therefore, Krell et al. [3] underlined the importance of developing scientific reasoning competencies during teacher education. Krell et al. defined scientific reasoning competencies as complex construct comprising three knowledge types for problem-solving (*knowing that*, *knowing how*, *knowing why*, cf. [4]) that are applied in cognitive processes (e.g., encoding, strategy development, cf. [1]). For this application, scientific reasoning skills such as formulating questions, testing hypotheses, planning and performing investigations, analyzing information systematically, and drawing reasonable conclusions from specific observations are required [1,5,6]. With regard to a rather broad understanding of scientific reasoning competencies, they can also be considered important for teachers to monitor and improve instructional quality in the science classroom. In such a context of instructional diagnosis, a teacher explicitly and systematically compares different characteristics of instruction in a data-based manner in order to be able to make appropriate instructional decisions [7]. Scientific reasoning competencies in such a context of diagnosis can be considered similar to conceptualizations of diagnostic competences [8,9]. For diagnostic competences, different components such as professional knowledge and diagnostic activities have been distinguished. Knowledge has been conceptualized in terms of the content-related facets pedagogical content knowledge (PCK), content knowledge (CK), and psychological-pedagogical knowledge (PK) as well as with regard to types of knowledge (*knowing that, knowing how, knowing when and why*) [10]. Since knowledge is applied in diagnostic contexts, and since within those contexts, the diagnostic focus can vary, all content-related knowledge facets may be of importance. Therefore, *knowing that, knowing how, knowing when and why* is not restricted to CK, but can also be distinguished for PCK and PK, which can be seen as an extension of the definition given in the scientific reasoning competencies approach. The conceptualization of diagnostic activities in which knowledge is applied in order to solve specific problems can be seen as equivalent to scientific reasoning skills [9].

Whereas many studies in recent years have examined the structure of professional knowledge and developed instruments to measure different knowledge facets (e.g., COACTIV: *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers* [11]; ProwiN: *Professional Knowledge of Teachers in Science*, [12]), there is still a lack on how to effectively promote the content-related knowledge facets PCK, CK, and PK and whether there are differences regarding the effectiveness of training procedures. Additionally, the effects of fostering professional knowledge on the execution of scientific reasoning skills are not well studied either. Within a biology context, the present study addresses this issue by investigating whether pre-service biology teachers' professional knowledge can be supported by self-directed knowledge acquisition via texts and if this support affects their scientific reasoning skills about instruction expressed as diagnostic activities.

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

1.1 Conceptualizing Teachers' Professional Competence in terms of Diagnosing

Teachers' professional competence has been studied from a cognitive perspective focusing dominantly on teachers' knowledge facets (e.g., [11,13,14]) and from a situated perspective including the context in which instructional decisions have to be made (e.g., [15–17]). The competence as a continuum model combines both perspectives and defines professional competence as a continuum with different components spanning from a teachers' dispositions (e.g., professional knowledge, beliefs) that underlie situation-specific skills, which in turn inform the teacher's actual instruction [18]. Dispositions are defined as "underlying characteristic of a person" [19] (p. 97) that can be regarded as cognitive in terms of professional knowledge and as affect-motivation in terms of teachers' belief, interest, or motivation [18,20]. Conceptualizations of situation-specific skills refer to teachers' adequate coping with teaching situations and allow an action-orientated assessment of variables that take situated learning approaches into account [20,21]. In the context of reasoning, such skills can be operationalized as scientific reasoning skills [22]. Furthermore, teachers' instruction in the classroom is considered to be decisive for teaching effectiveness. Effective teaching can be described by generic and subject-specific dimensions of instructional quality, which influence students' outcomes [23–25]. As part of their professional competence, teachers should know about these dimensions and included instructional quality features in order to reason about them, to diagnose instructional processes, make appropriate decisions, and adapt teaching (cf., [17,26]). Within these situation-specific processes, teachers evaluate data (e.g., from monitoring scientific inquiry steps, the elicitation of student thinking, the use of three-dimensional models) to inform their pedagogical reasoning and decision-making [27]. Since such processes are part of the systematic and continuous generation and evaluation of knowledge about students and instructional characteristics, they can be summarized as diagnosing, which counts as an important dimension of teachers' professional competence [28,29]. Taking the data-based process into account, diagnosing can be considered as a type of scientific reasoning including several epistemic activities teachers can make use of [30]. In the context of diagnosis, these activities have also been described as diagnostic activities [9]. Therefore, teacher education should not only foster teachers' professional knowledge regarding instructional quality features but also enable them to "apply their knowledge in diagnostic activities according to professional standards to collect and interpret data in order to make decisions of high quality" [9] (p. 9). Thus, diagnostic activities represent scientific reasoning skills used to specify situation-specific skills in the context of diagnosis.

Depending on the context in which knowledge has to be applied, different content-related knowledge facets and types of knowledge may be critical [10]. In addition, different diagnostic activities have been described, but not all of them are considered relevant for diagnosing specific situations [9] since in some situations, it might be sufficient to generate and evaluate evidence, whereas in other situations, the generation of hypothesis or the creation or redesign of artifacts may be more important for knowledge generation. In the following, different approaches to conceptualize teacher professional knowledge and skills are described.

1.1.1 Conceptualizing Teachers' Professional Knowledge

Effective biology teaching requires different knowledge types. Förtsch et al. [10] distinguished knowledge related to facts, terms, and principles (mostly referred to as declarative knowledge or *knowing that*), and action-related knowledge (*knowing how, knowing when and why*). *Knowing that* means, for example, that a teacher can correctly list the advantages and disadvantages of a specific model. *Knowing how* refers to knowledge about actions, procedures, or manipulations, and is applied, for example, when a teacher deals with students' ideas. *Knowing when and why* relates to knowledge about when and why to apply particular procedures to achieve particular goals, for example, knowing when and why students' errors within a certain topic are dealt with [10,31].

In addition to the division into different types of knowledge, content-related knowledge facets can be classified. Based on Shulman's classification [32,33], most models focused on PCK, CK, and PK that build the core of the construct [11]. Knowledge can be described for all three knowledge facets in terms of declarative and action-related knowledge [10]. The subject-independent facet PK contains knowledge that counts as necessary for classroom management, classroom assessment, and organization to facilitate an effective learning atmosphere in which pedagogical strategies can be applied [34,35]. Broader conceptualizations as used in the BilWiss project include the PK-dimensions instruction (which is further divided into the sub-dimensions instructional quality features and teaching strategies/methods), learning and development, diagnostics and evaluation, educational theory, school as an educational institution, and teaching as a profession [36]. In general, the conditions a science teacher establishes in the classroom are assumed to provide the basis on which PCK and CK can be used [34]. Since PCK and CK count as subject-specific, they are most important for science education. CK describes the knowledge of subject matter, discipline-specific methods for generating knowledge, and the conceptual understanding of specific topics, which researchers emphasized as a necessary but insufficient precondition for the development of PCK [11,33,37]. With regard to a specific subject matter, PCK includes subject-specific knowledge about corresponding (mis)conceptions of particular students, knowledge about subject-specific structures of instruction, and corresponding teaching strategies, and was shown to be highly predictive for instructional quality and students' achievement [17,21,38]. However, researchers emphasized that it is not only this stage of knowledge that forms PCK but also the knowledge that is closely related to the actual practice and thus is more dynamic [38–40]. Science teaching includes taking students' prior learning into account, facilitating linkages between concepts, or choosing and utilizing instructional strategies that best suit particular teaching moments. Such tendencies underpin teachers' pedagogical reasoning, which is the heart of teaching, and are regarded as components of PCK rather than equivalent to PCK (see [41,42]).

For measuring professional knowledge, researchers mostly used paper-pencil tests with validated test items [11,43]. However, to measure context-dependent, practice-oriented knowledge, other approaches than paper-pencil assessments are needed that take the situated character of knowledge application into account [44].

1.1.2 Conceptualizing Situation-specific Skills for Reasoning about Instruction

Researchers assume that teachers' professional knowledge underlies their situation-specific skills that teachers use to systematically solve specific situations in the classroom or to inform subject-specific instruction [18]. When solving (problematic) situations, teachers engage in reasoning processes in order to make decisions. From a scientific stance, "scientific reasoning encompasses the reasoning and problem-solving skills involved in generating, testing and revising hypotheses or theories, and in the case of fully developed skills, reflecting on the process of knowledge acquisition and knowledge change that results from such inquiry activities" [1] (p. 61). In this definition, it becomes evident that scientific reasoning comprises specific processes that aim at generating knowledge [1,30]. The mentioned processes are in line with Nowak et al. [6], who identified three main processes that are central to scientific reasoning: (1) asking questions and formulating hypotheses, (2) planning and performing an investigation, (3) analyzing data and reflecting on the investigation. For effective engagement in any of these reasoning processes, scientific reasoning skills including reasoning from evidence are required [22,45]. Several skills have been described either with regard to the three main processes (cf. [6]) or with regard to epistemic activities that have been found relevant for generating knowledge in different domains [30]. Overall, eight epistemic activities have been described: (1) *problem identification*, (2) *questioning*, (3) *hypothesis generation*, (4) *construction and redesign of artefacts*, (5) *evidence generation*, (6) *evidence evaluation*, (7) *drawing conclusions*, and (8) *communicating and scrutinizing*. Such scientific reasoning skills are considered vital not only for teachers' classroom instruction [46] but for every human's understanding of the world and for the development of responsible citizenship [1,47]. Accordingly, it is important that science teachers master such skills: Firstly, to promote them among their students and, secondly, to apply them in the context of reasoning processes about science or biology instruction.

Within the broader understanding of scientific reasoning, applying scientific reasoning on instruction can be seen as an evidence-based process of systematically collecting data, generating and evaluating evidence, and drawing inferences in order to produce a diagnosis and make instructional decisions (cf. [9,27,48]). Therein, evidence is not only the product of an experimental investigation but can also consist of statements describing observations (cf. [45]). Overall, scientific evidence-based reasoning in the context of diagnosing enables teachers to diagnose their students and also instructional features that are important in terms of instructional quality, and thus, for student learning (cf. [11,49]). In teacher education, pre-service teachers should therefore also develop knowledge and skills to enact scientific reasoning in different contexts, such as diagnosing, for which specific tools are needed [3].

In the context of diagnosing, scientific reasoning skills have been operationalized as diagnostic activities. Diagnostic activities describe those activities that teachers execute for data-evaluation within situation-specific diagnostic contexts and that are more clearly observable than solely cognitive processes that underlie diagnosing [9,50]. Diagnostic activities have been mentioned in several studies (e.g., [51,52]), but an explicit definition of different activities is mostly missing. A more differentiated approach was recently made by Heitzmann et al. [9], who translated the eight epistemic activities introduced by Fischer et al. [30] into eight diagnostic activities relevant for the goal-oriented process of diagnosing. In the following, these diagnostic activities are illustrated with examples from teaching:

- *identifying problems* (e.g., a teacher recognizes a noteworthy incident in classroom instruction driven by prior knowledge, cf. [53]),
- *questioning* (e.g., a teacher asks for reasons of the identified problematic incident),
- *generating hypothesis* (e.g., a teacher makes an assumption about the underlying problem of the teaching situation),
- *constructing artifacts* (e.g., a teacher generates tests/tasks to be used for (further) data collection),
- *generating evidence* (e.g., a teacher/observer uses the test/task or systematically observes and describes the situation, for example, with regard to relevant student or teacher behavior),
- *evaluating evidence* (e.g., a teacher interprets data and evaluates the extent to which it supports a demanded standard),
- *drawing conclusions* (e.g., a teacher derives [behavioral] consequences from the evaluation of multiple data sources,
- *communicating the process/results* (e.g., a teacher shares findings and feedback can be given; afterward, further measures can be taken or alternative instructional strategies can be implemented).

The eight diagnostic activities can be understood as a reservoir of activities teachers can use for diagnosis. Which diagnostic activities are appropriate may differ with regard to the discipline or the diagnostic focus [9,54]. In studies that investigated diagnostic activities in the context of teacher education, the diagnostic activities *generating hypothesis*, *generating evidence*, *evaluating evidence*, and *drawing conclusions* were considered to be particularly relevant [48,54,55]. Furthermore, it is noteworthy that some diagnostic activities show similarities with conceptualizations of situation-specific cognitive skills such as perceiving, interpreting, and decision-making (PID model, [56]) or professional vision [57,58]. Both conceptualizations have been used in the context of video analysis and the diagnosis of classroom instruction. From the perspective of the PID-model, teachers' abilities to *perceive* particular events in instructional settings, to *interpret* the events, and to *make decisions* either as anticipating answers to student ideas or proposing alternative teaching strategies have been identified as crucial in terms of professional competence. From the perspective of professional vision, the skills of *noticing* (paying attention to noteworthy events) and *reasoning* (describing noteworthy events, explaining by linking pedagogical concepts and principles to observed events, and predicting possible consequences as specification of teachers' decision making) are highlighted [57]. Therefore, researchers using diagnostic activities in the context of video analysis have operationalized *generating evidence*, *evaluating evidence*, and *drawing conclusions* as reasonable diagnostic activities for assessment designs [55].

Even when scientific reasoning skills, and more explicit the epistemic purpose underlying a diagnostic activity, are transferable across disciplines [59], to a certain extent, the application of scientific reasoning skills is discipline- and context-specific as well as knowledge-dependent (cf. [60,61]). Therefore, the adequate execution of diagnostic activities in the context of diagnosing biology instruction may rely on subject-specific facets of professional knowledge such as teachers' PCK (cf. [8,62]).

1.2 Fostering Professional Knowledge and (Scientific) Reasoning Skills

Due to the complex interaction and interdependence of professional knowledge and scientific reasoning skills, such as diagnostic activities, pre-service teachers need varying opportunities to develop knowledge and apply diagnostic activities during their teacher education. The common division into three content-related knowledge facets is also reflected in the university organization, in which the knowledge facets are taught in separate courses and lectures [63], while different knowledge types are addressed more or less explicitly across all courses. Standard working methods in higher education include text-based procedures requiring pre-service teacher to acquire knowledge self-directed, lecture-based procedures in which specific information is presented by a lecturer, mixed forms of text- and lecture-based instruction, and situated approaches to learning that represent scenarios from real-world demands, for example, in video vignettes (cf. [64,65]). Depending on the context, on learners' prerequisites, on structure and content of materials, and specific learning goals, the effects of instructional support on knowledge acquisition may vary.

Barth et al. [66] compared the effects of self-directed knowledge acquisition and direct instruction of knowledge about classroom disruptions that represent an aspect of PK on three cognitive outcomes: (declarative) knowledge on classroom management, noticing critical incidents in the classroom, and knowledge-based reasoning. Results showed that direct instruction that was conducted by a university teacher and included a systematic introduction to the relevant content led to higher gains in knowledge on classroom management (PK) and improved the "ability to apply this knowledge in a simulated teaching situation (the video) through knowledge-based reasoning" [66] (p. 8). However, noticing (that corresponds to the diagnostic activity *problem identification*) was not affected. In addition, the self-directed acquisition of knowledge did not result in any significant effects. Kleickmann et al. [67] investigated conditions that are necessary for developing PCK in mathematics education. Experimentally manipulated treatments received instructions by an experienced lecturer on different combinations and sequencing of declarative and action-related PCK, CK, or PK over two days. Additionally, important content was repeated, and the participating pre-service teachers received handouts and had to carry out different tasks such as answering short questions or writing assignments to recapitulate the major contents of the instruction. Regarding direct instruction, their results showed that "explicitly addressing the knowledge of students, learning and teaching in concrete content domains, whether with or without antecedent CK instruction, appeared to be the most effective pathway" [67] (p. 126). A reanalysis of the data also underlined that instruction on PCK has small effects on CK or PK development as well [63]. The authors of the study attributed this to comparisons and reflections stimulated by the test questions that might have prompted particular aspects of CK and PK. Furthermore, Smit et al. [68] investigated the relationships between PCK, CK, and scientific inquiry attitudes. They found gains for both declarative PCK and CK measured with test items within a training program on scientific inquiry. To ensure all participants had a similar level of knowledge, a teacher educator gave a theoretical PCK input on scientific inquiry. Furthermore, inquiry-related videos were discussed. Results showed a major relationship between PCK and scientific inquiry attitudes. In addition, the input proved to be effective for PCK and scientific inquiry attitudes. The ensuing intervention consisted of a peer-coaching on lesson planning focusing on scientific inquiry skills. However, lesson planning was not found to affect professional knowledge. Possible effects on scientific inquiry skills have not been investigated.

Besides studies focusing on the three knowledge facets PCK, CK, or PK, training interventions that aim to improve situation-specific skills exist. Positive effects of university courses were shown, in which noticing and knowledge-based reasoning were fostered by using videos [57,69,70]. In most of these courses, pre-service teachers discussed and reflected teaching performances shown in video clips, which potentially addressed the diagnostic activities *problem identification*, *generating* and *evaluating evidence*, and *drawing conclusions*. To our knowledge, effects of instructional support via texts explicitly on activities relevant within diagnostic contexts have not been investigated yet.

However, a first approach to differentiate between measurements of cognitive dispositions in terms of knowledge facets and situated skills was recently made by Gess-Newsome et al. [40]. As part of a three-year professional development training, participants studied curriculum materials, discussed issues of effective pedagogy, and deepened CK to promote different facets of knowledge and skills. Investigated facets of teachers' knowledge and skills were declarative CK, general PK, two separate PCK-constructs (PCK-PK and PCK-CK) as well as inquiry-oriented teacher practice. Even though general PK was conceptualized as a cognitive knowledge facet in their initial conception, the authors finally recognized that by using an observation protocol to assess PK in video-recorded classroom sessions, they actually measured a skill instead of declarative PK. Furthermore, the assessment of the PCK-constructs was situated and can be considered as an approach to elicit skills as enacted form of PCK (cf. [41]). These PCK-related skills reflected the application of reasoning skills in terms of diagnostic activities on a meta-level. Skills referred to the abilities to describe a lesson (i.e., *generating evidence*), to explain rationales for instruction (i.e., *evaluating evidence*), and to make instructional decisions (i.e., *drawing conclusions*). In addition, Gess-Newsome et al. [40] investigated the development of teachers' inquiry-based instruction and, thus, included a second measure related to reasoning skills directly used in action. As the result of the three-year professional development training, the authors found an increase of all investigated facets of teachers' knowledge and skills, indicating the effectiveness of integrating multiple pathways of teachers' professional learning.

Despite the number of studies carried out in the field of professional competence training, only a few studies can be found that explicitly investigated a certain type of knowledge acquisition (self-directed via texts, lecture or instruction, video-club) and its impact on the knowledge facets PCK, CK, and PK, as well as on teachers' skills such as diagnostic activities in a systematic way. Previous studies have rarely focused on the investigation and support of all knowledge facets equally. Moreover, in some studies, the conceptualization of investigated variables lacks preciseness or has not been clearly considered.

Assessment situations should include the explicit measurement of all knowledge facets and types or at least situate measures to specific facets and types for a more fine-grained differentiation and analysis of professional knowledge in order to examine knowledge development in relation to methodological approaches and training measures (cf. [10,71]). This is important for understanding the nature of the individual knowledge facets and skills, and for clarifying whether and what type of intervention is effective. It also provides important information for practical implementation.

1.3 Motivation of the Study and Research Questions

Teacher education at universities continues to provide a great deal of knowledge acquisition about text-based instruction, although it is unclear to what extent this is effective. The question arises whether this kind of learning setting is best for pre-service teachers with little classroom experience. Therefore, our first goal was to investigate the effects of knowledge acquisition via texts and how self-directed knowledge acquisition via texts affects pre-service teachers' professional knowledge facets PCK, CK, and PK. Our second goal was to analyze if knowledge acquisition also affects the application of scientific reasoning skills within the context of diagnosing instructional quality. Since it is assumed that the use of scientific reasoning skills, such as diagnostic activities, relies to some degree on an individual's knowledge (cf. [61]), there might not only be an effect of self-directed knowledge acquisition via texts on the professional knowledge facets but on diagnostic activities as well [17,41]. Thus, the present study addresses the following research questions:

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

2. Materials and Methods

2.1 Design and Sample

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

The experimental design of the study contained a pre-test (day 1), a post-test (day 3) and featured five different treatments (intervention on day 2). Pre- and post-data were collected in two steps each (see Figure 1). First, pre-service biology teachers completed three paper-pencil tests to measure their PCK, CK, and PK, which were the same in pre and post-test. Second, we used the video-based assessment tool DiKoBi *Assess* to measure pre-service teachers' diagnostic activities pre (DiKoBi I *Assess*) and post (DiKoBi II *Assess*). The assessment tool provides videotaped classroom situations that have to be diagnosed with regard to different subject-specific dimensions (see section 2.3.2 *Video-based assessment tool*). The diagnostic tasks were the same for DiKoBi I and II Assess; both versions differed only in the content of the classroom situations shown. Both pre- and post-measurements took 120 min each. The intervention lasted 90 min and consisted of five different treatments, in which information on either (1) PCK, (2) CK, (3) PK, (4) a combination of these three knowledge facets, or (5) none of these knowledge facets (control group) was acquired in a self-directed way. The intervention covered declarative and action-related knowledge relevant for teaching the topic "skin". The same topic was also addressed in the videotaped classroom situations of DiKoBi *Assess* [72].

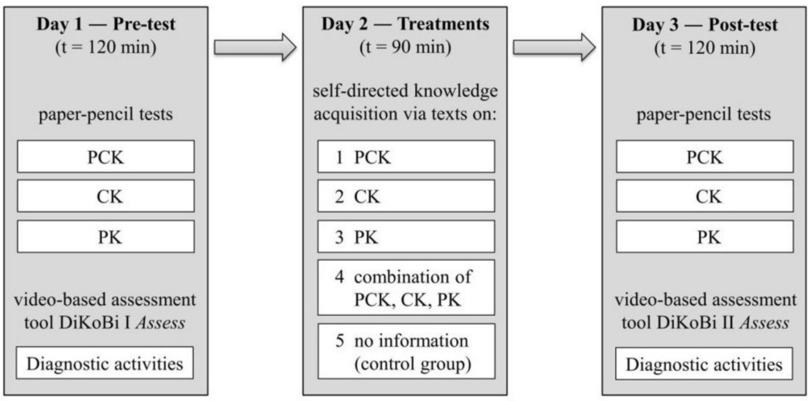

Figure 1. Design of the study

The sample consisted of 81 pre-service biology teachers (75.3 % female; average study semester: M = 3.9, SD = 1.3; age in years: M = 23.6, SD = 3.9). 48.1 % of the pre-service teachers attended the academic track of teacher education, qualifying them for future teaching at German secondary schools ("*Gymnasium*"); 51.9 % attended programs for the non-academic track that prepares students for a vocational career (for an overview of the German school system see [73]).

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

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

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

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

2.2 Description of the Treatments

After reviewing the literature on generic and subject-specific features of instructional quality and including theoretical aspects that were relevant for pre-service teachers' reflection of classroom situations in DiKoBi *Assess*, content-specific aspects regarding PCK, CK, and PK were identified and summarized in texts (for an overview about the content included in the texts, see Appendix A). The texts either contained information on only one knowledge facet (treatments 1-3, see Figure 1) or represented a combined form of all three knowledge facets (treatment 4). A control group (treatment 5) did not receive any information according to the three knowledge facets. After a five minute lasting introduction into day 2 of the study, the participants of each treatment worked individually for 85 minutes on the associated texts. In treatments 1-4, participants were guided by identical tasks that were adapted with regard to the specific content of the texts (for an example, see Appendix B). The participants were asked to highlight important information in the texts (task 1), to supplement an already outlined concept map on the basis of the highlighted information (task 2), and to apply this information by evaluating a statement or situation and providing alternatives (task 3). The tasks were constructed to consider declarative and action-related knowledge that were also part of the professional knowledge tests used [43]. However, the major share relates to declarative knowledge.

2.3 Measuring Professional Knowledge and Diagnostic Activities

2.3.1 Professional Knowledge Tests

Three paper-pencil tests for measuring PCK, CK, and PK were utilized. The tests included open-ended items (required a written response in a text field), single best answer (SBA) items (required the selection of one correct answer from a set of possible responses consisting of multiple distractors and one correct answer), and multiple true/false items (all of the possible responses had to be assessed for their validity) [74].

The PCK and CK test covered declarative and action-related knowledge about the topic "skin" (in accordance with the topic covered in DiKoBi *Assess* and the intervention). For example, declarative knowledge (*knowing that*) for PCK was addressed by asking for advantages and disadvantages to a specific model, which shows the structure of the human skin. Action-related knowledge in terms of *knowing when and why* was measured by asking for possible reasons students develop specific misconceptions on a specific biology topic after the learning process (cf. [43]). Both the PCK and CK test were adapted versions of the professional knowledge tests used in the ProwiN project [43,75]. The PCK test included eight open-ended items and five SBA items. Therefore, we assumed to elicit pre-service biology teachers PCK with the different items since for responding, the pre-service teachers had to draw on their individual specialized knowledge. The PCK test covered two important components of biology teachers' PCK: knowledge of instructional strategies (model use and use of experiments) and knowledge of students' errors [12]. The CK test included 13 open-ended items and 15 SBA-items. Criteria for item scoring of both the PCK and CK test were provided in two separate coding manuals. Precise descriptions of the scoring process can be found in Kramer et al. [8]. To ensure objective and reliable coding, ten percent of both the PCK and CK tests were coded by two independent raters utilizing the coding manuals. A high agreement between the two raters has been shown by the results of two-way random intra-class correlations ($ICC_{absolute}$): PCK: $ICC_{absolute}(310,310) = 0.84$, $p < 0.001$; CK: $ICC_{absolute}(341,341) = 0.97$, $p < 0.001$ [76].

For assessing PK, we used a short, adapted version of a paper-pencil test utilized in the BilWiss project covering the dimension *instruction* [36,77]. This dimension of the PK test referred to basic dimensions of instructional quality containing declarative items about generic features such as *classroom management*, *supportive climate*, and *cognitive activation* [11,24,78], as well as items on general pedagogical issues of teaching such as teaching methods. Since the differentiation between generic and subject-specific features of instructional quality was an important element of the video-based assessment tool DiKoBi *Assess*, the dimension *instruction* was best suited to our construct. For PK-measurement, participants had to answer five SBA-items and ten multiple true/false items. Item scoring followed the instructions from BilWiss [36,77]. Precise descriptions of the scoring process can be found in Kramer et al. [8].

Afterward, each knowledge test was evaluated using the Rasch partial credit model (PCM), which resulted in PCK, CK, and PK Rasch person measures for each respondent for each test instrument [79,80]. For evaluating data fit, we utilized item Outfit-MNSQ (mean-square) values, item reliability and person reliability for each test. A productive measurement is shown by item Outfit-MNSQ values below 1.5 [81]. If item reliability is high, both the range of item difficulty and the sample size can be considered as appropriate to measure the variables precisely. The person reliability is a measure of internal consistency. Person reliability is impacted by the length of the test and the range of abilities of respondents [82]. Item fit statistics of the PCK, CK, and PK test showed good fit values (see Table 2). To compare data from the identical pre- and post-tests, we anchored items from the pre-test with appropriate items from the post-test. After analyzing pre- and post-test of each knowledge facet utilizing Differential Item Functioning [80], we included ten anchor items for the PCK test, 23 anchor items for the CK test, and 11 anchor items for the PK test. Those items, which produced a measurement bias for pre- and post-test were excluded from anchoring.

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

Knowledge facet	Number of items	All item Outfit-MNSQ	Person reliability	Item reliability
PCK $pre_{anchored}$	13	< 1.36	0.65	0.91
PCK post	13	< 1.10	0.62	0.93
CK $pre_{anchored}$	28	< 1.37	0.70	0.95
CK post	28	< 1.36	0.76	0.96
PK $pre_{anchored}$	15	< 1.28	0.54	0.95
PK post	15	< 1.18	0.60	0.95

2.3.2 Video-based Assessment Tool DiKoBi *Assess*

To measure the three diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*, which are applicable for diagnosing instructional quality [55], we used the video-based assessment tool DiKoBi *Assess* that is embedded in an online survey platform [83]. DiKoBi contains short staged video clips showing challenging biology classroom situations on the topic "skin". DiKoBi *Assess* consists of six videotaped classroom situations, which represent one whole biology lesson and refer each to another subject-specific dimension of instructional quality that was found to be empirically effective for student achievement in science instruction [84]: (1) *level of students' cognitive activities and creation of situational interest*, (2) *dealing with (specific) student ideas and errors*, (3) *use of technical language*, (4) *use of experiments*, (5) *use of models*, (6) *conceptual instruction*. The evaluation of these dimensions is applicable to any biology lesson regardless of the specific content to be taught and refers to the identification of subject-specific instructional quality features [72].

For each of the six classroom situations of DiKoBi *Assess*, pre-service teachers had to identify challenging aspects of the shown situation of biology instruction and reason about them by describing the identified challenging aspects (diagnostic activity *generating evidence*), by explaining why there is room for instructional improvement (including theoretical references) (diagnostic activity *evaluating evidence*), and by proposing alternative teaching strategies (diagnostic activity *drawing conclusions*) (cf. [54,56,57]). Pre-service teachers' diagnostic activities were measured in an open-ended format with short-answer items [85]. For scoring, written answers were compared with predefined sample solutions of content-related coding variables that have been compiled with regard to the literature and research results on the subject-specific dimensions of instructional quality. The content-related coding variables referred to subject-specific instructional quality features that represent challenging aspects of biology instruction. For example, when using experiments, the steps of scientific inquiry should be implemented; or when using three-dimensional models, the models should be applied as scientific inquiry instruments including testing the models or judging the purpose of the models (cf. [3,5,72]). Depending on the shown situation of biology instruction, participants' identified challenging aspects had to refer to corresponding features of subject-specific instructional quality. Therefore, diagnostic activities were measured content-dependently. In total, *generating evidence* was measured with 13 coding variables, *evaluating evidence* was measured with six coding variables, and *drawing conclusions* with 12 coding variables. Depending on the quality of the executed diagnostic activity (e.g., observable in not accurate/vague answers, or in more detailed, elaborated answers), Zero (0), 1, or 2 points were used for scoring answers corresponding to the diagnostic activity *generating evidence*, 0, 1, 2, or 3 points for scoring answers corresponding to *evaluating evidence*, and 0, 1, or 2 points for answers corresponding to *drawing conclusions* (cf. [8]). For a high-quality scoring of answers corresponding to the diagnostic activity *generating evidence*, it was important that the provided answer contained a systematic description of an observed challenging instructional aspect. A high-scored answer of the diagnostic activity *evaluating evidence* contained references to scientific concepts that were used to justify the claim that was made. High-quality in *drawing conclusions* became visible by specifically described alternative strategies that were derived based on the preceding steps of scientific reasoning. Results of several qualitative validation steps showed that the staged challenging instructional aspects are received as authentic, that they could be sufficiently identified by in-service biology teachers and that the assigned tasks validly measured the assumed diagnostic activities. Further information regarding the assessment tool including validation results can be found in Kramer et al. [55,72].

For this study, we used two versions of DiKoBi *Assess*, which differed in the specific sub-theme of the embedded videos on the topic "skin":

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

After data collection, Rasch PCM was used to analyze person measures for executing all three diagnostic activities in six classroom situations. This was done separately for pre- and post-test, since the corresponding versions of DiKoBi contained different sub-themes that may have had an impact on the execution of diagnostic activities. Fit statistics showed good fit values for pre- and post-test including a one-dimensional construct of diagnostic activities (diagnostic activities pre/post: 31/31 items, all item Outfit-MNSQ < 1.32/1.38, person reliability = 0.77/0.73, item reliability = 0.90/0.85; note: for diagnostic activities pre, one item in DiKoBi I *Assess* produced inestimable high values). The one-dimensional construct was used due to weak reliability values when calculating person measures separately for the three diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*.

Our video-based approach follows Kersting et al. [17], who "use video clips of authentic classroom events as prompts to elicit teachers' analyses, which are in turn assumed to draw on teachers' knowledge" (p. 571). Therefore, we assume that the videotaped classroom situations elicit pre-service biology teachers' diagnostic activities for reasoning about the biology-specific challenges and features in both versions of the assessment tool, which are in turn assumed to rely on pre-service teachers' declarative and action-related PCK (cf. [8]).

2.4 Data Analysis

First, measures of all variables [pre/post: PCK, CK, PK, diagnostic activities] were separately analyzed using the Rasch PCM [79] with the software Winsteps 3.81 [86]. Second, Pearson's correlations and descriptive results were calculated utilizing the softwares IBM SPSS Statistics (version 26) and Microsoft Excel (2010) to describe the development and intercorrelation between all variables relevant for this study. The main analysis was done in two steps. (I) To answer RQ1a and RQ1b, we ran mixed ANOVAs for PCK, CK, and PK to analyze the main and interaction effects between time and treatments. (II) To answer RQ2, we chose an analysis of covariance (ANCOVA) to analyze effects on diagnostic activities in post-test while controlling for diagnostic activities in pre-test. There was homogeneity of the error variances of all variables, as assessed by Levene's test ($p > 0.05$), as well as homogeneity of covariance, as assessed by Box's test [(personal) PCK: $p = 0.53$; CK: $p = 0.57$; PK: $p = 0.53$) and homogeneity of regression slopes (diagnostic activities: $p = 0.20$).

Table 3. Intercorrelations, including mean and standard deviation. Mean values of PCK, CK, and diagnostic activities are negative, because test items were rather difficult for pre-service biology teachers.

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

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

3. Results

A descriptive overview about means and standard deviations, as well as intercorrelations between all variables, is given in Table 3. The mean values represent the average person ability from the PCM of the corresponding variables. Intercorrelations emphasized the importance of PCK since there was a low to moderate correlation between PCK and most of the pre or post-measured variables (Cohen, 1988). For example, PCK pre was significantly correlated with PCK post ($r = 0.57$, $p < 0.001$), CK post ($r = 0.30$, $p = 0.006$), PK pre ($r = 0.22$, $p = 0.048$), and diagnostic activities pre ($r = 0.33$, $p = 0.002$). Furthermore, the descriptive results showed that pre-service teachers' PCK, CK, and PK increased from pre to post (see Table 3). Note that the average person measures of diagnostic activities pre and post are not directly comparable because they have not been anchored, as we used two separate measurement instruments in pre and posttest pre (pre: DiKoBi I *Assess* and post: DiKoBi II *Assess*). However, a descriptive comparison between diagnostic activities pre and diagnostic activities post can be made based on the quality of the diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*. Table 4 shows how often each diagnostic activity was scored with 0, 1, 2, or 3 points. It is noteworthy that the total scores of *generating evidence* and *drawing conclusions* decreased from pre to post, while the total score of *evaluating evidence* increased. Furthermore, the dispersion increased from pre to post for the diagnostic activities *generating* and *evaluating evidence*. More often 0 points (low quality of diagnostic activity) but also 2 or 3 points (improved quality of diagnostic activity) were used for scoring. In contrast, fewer answers were scored with 1 point in DiKoBi II *Assess*. For the diagnostic activity *drawing conclusions*, frequency of 0 points increased whereas frequency of 1 and 2 points decreased, indicating an overall decrease of quality.

Table 4. Absolute frequency of points that were used for scoring the quality of the three diagnostic activities. *N* refers to the total number of activities scored. This number is calculated from the number of participants (81) and the number of content-related coding variables for the respective diagnostic activity and is therefore the same for pre and post.

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

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

Next, we ran mixed ANOVAs for PCK, CK, and PK to analyze time effect, treatment effect, and interaction effect between time and treatment (RQ1a, RQ1b). Main effects of time were found for both PCK and CK. They confirmed the positive descriptive trend of knowledge acquisition since a statistically significant increase of mean person abilities of PCK and CK was measured from pre to post. However, there was no statistically significant increase from pre- to post-measurement for PK. Furthermore, there was no statistically significant main effect of treatment or interaction between time and treatment for any knowledge facet when each treatment was considered individually (see Table 5).

Table 5. Results of mixed ANOVAs of the knowledge facets

	df	F	p	partial η^2
PCK				
- time effect	**1,76**	**118.28**	**< .001**	**0.61**
- treatment effect	4,76	0.53	0.715	0.03
- interaction effect	4,76	1.11	0.359	0.06
CK				
- time effect	**1,76**	**31.60**	**< .001**	**0.29**
- treatment effect	4,76	0.15	0.963	0.01
- interaction effect	4,76	2.13	0.085	0.10
PK				
- time effect	1,76	3.60	0.062	0.05
- treatment effect	4,76	0.17	0.952	0.01
- interaction effect	4,76	0.72	0.583	0.04

Note: Significant results are highlighted in bold.

However, since the maximum number of participants per treatment did not exceed 18, we merged treatments that included CK in a second step. This step was done because the p-value of 0.085 of the interaction effect between time and treatment for the CK group was considerably lower compared to PCK or PK. This p-value might indicate a possible underlying effect of the self-directed knowledge acquisition via texts in terms of CK acquisition that might not have been detectable due to the small number of participants per treatment. Since it was not possible to increase the number of participants in the overall sample for the time of the study, a fallback solution was applied: treatments containing CK and treatments not-containing CK were merged (group 1: CK-treatment and combination-treatment; group 2: PCK-treatment, PK-treatment, control group). Since our PCK-treatment did not contain CK content, we assigned the PCK-treatment to not-containing CK (group 2). Accordingly, the PCK test was constructed in such a way that PCK could be measured as independently as possible from CK (cf. [49,75]).

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

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

4. Discussion

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

In summary, neither PCK, CK, PK nor diagnostic activities were significantly affected by any of the study treatments. The knowledge facets PCK and CK have significantly increased from pre to post; however, this increase could not be explained by the self-directed knowledge acquisition (RQ1a, RQ1b, RQ2). This finding is in line with other research showing, for example, that PK could not effectively be fostered through self-directed knowledge acquisition compared to direct instruction [66]. Therefore, the methodological approach to knowledge acquisition via texts as it was implemented in the present study has not proven to be effective. However, that does not mean that self-directed knowledge acquisition via texts that represents one common learning practice at German universities is generally ineffective. Effectiveness may depend on the specific actions that are initiated by the instructional approaches. For example, Kyriakides et al. [87] found that both direct instruction and self-directed constructivist approaches can benefit student outcomes, depending on what exactly the teacher and the students do during instruction. Therefore, even though we made an effort to increase pre-service teachers' engagement with the learning material, the utilized tasks may not have been activating enough or appropriate to promote in-depth learning. Moreover, possible small effects of the self-directed knowledge acquisition via texts might have been overlaid by other effects, possibly resulting from the video-based work. However, descriptive comparisons regarding the quality of the diagnostic activities pre and post showed an increase in terms of quality that was slightly noticeable for the diagnostic activity *generating evidence* and particularly noteworthy for *evaluating evidence*. Whereas pre-service teachers' evidence evaluations in DiKoBi I *Assess* were often superficial and vague, their quality slightly increased in DiKoBi II *Assess* in terms of more frequent concept references and explicit linking of observations and theoretical references. This primarily indicates a potential impact of the self-directed knowledge acquisition on the diagnostic activity *evaluating evidence*. Similar findings regarding the relationship between knowledge and interpretive processes have also been described in other studies (cf. [88]). The findings on the decrease in the quality of the diagnostic activity *drawing conclusions* from pre to post might indicate that other approaches of instructional support are necessary for the promotion of this activity, which cannot be provided via text-based instruction (that included a large amount of declarative knowledge). In addition, affective-motivational aspects have to be taken into account, because in order to set up an alternative strategy, longer answers were required, for which the study participants might not have been motivated enough (cf. [89]). However, further research is needed to make more reliable conclusions and differentiate pathways of knowledge and skill development.

Hence, two other important questions remain whose answers contribute to the debate on measuring knowledge and skills of science teachers: First, how can the increase in knowledge in our study be explained, if not by the treatments? Second, which ways of knowledge acquisition might be more effective for university teaching? Regarding the first question, we want to refer to "the use of classroom video as a tool for bringing the central activities of teaching into the [professional development] setting" [90] (p. 1099). The greatest effects reported in this article are time effects of PCK-measures and, therefore, were independent of treatment. Researchers within science education already underlined that the use and prompted analysis of classroom performances challenges pre-service teachers' thinking and thus can activate their knowledge and make it accessible [17,40,69,91]. Our hypothesis is that the increase of PCK is due to the work with the video-based assessment tool DiKoBi *Assess*. The observation and diagnosis of biology-specific classroom situations may have elicited existing subject-specific knowledge. Working on the tasks in DiKoBi *Assess* and engaging in scientific evidence-based reasoning on biology instruction required pre-service teachers to apply diagnostic activities. This application of diagnostic activities to a specific situation of biology instruction may have contributed to the promotion of PCK. By observing and describing challenging aspects of biology instruction (i.e., *generating evidence* that helps encoding, cf. [1]), pre-service teachers directed their focus to very specific aspects. For the evaluation of this evidence (i.e. the observed and described challenging aspects) a linkage with broader principles they represent and thus the elicitation of professional knowledge had to take place [57,70]. The scientific reasoning skill *evaluating evidence*, in particular, is considered important in order to interpret classroom interactions and inform appropriate follow-up decisions [92]. Thus, implementing opportunities in which diagnostic activities can be applied or are even fostered may in turn have an impact on pre-service teachers' PCK. This assumption can be seen as an indication of the bidirectional relationship between knowledge and skills [93]. Consequently, the results of the study suggest that scientific reasoning about instructional quality can potentially promote PCK and that the application of skills, such as diagnostic activities, to video-based settings thus seems to be more suitable for knowledge development than instructional support via texts.

Building on these thoughts, we want to emphasize the relevance of using and reflecting on practical examples for the development of pre-service teachers' professional knowledge (cf. [91]). While a profound declarative knowledge base is considered important (especially in terms of CK) and may still be provided via specialized texts, PCK and PK are much more action-oriented and therefore require other forms of knowledge acquisition (cf. [94]). Our suggestion based on the present findings is to provide learning opportunities in which pre-service teachers engage in scientific reasoning about instructional quality, for example, via video-based tools. Other researchers have already made similar suggestions. König et al. [88], for example, used video vignettes to underline the importance of practical insights into teaching to improve teachers' general pedagogical knowledge. By setting the focus on practical scenarios, they also addressed reasoning skills as necessary components for the acquisition and transformation of knowledge. The use of video-based tools can be considered an appropriate approach to provide opportunities for the assessment and development of PCK and PK in teacher education that count as an important part of diagnostic competences [9,62,94].

Besides the use of video-based tools, other ways of supporting pre-service teachers' knowledge development have been discussed with regard to direct instruction. Effective approaches to knowledge acquisition often included an experienced lecturer in addition to text-based work in science teacher education. In contrast to static texts, a lecturer can "explicitly [address] the knowledge of students, learning and teaching in concrete content domains" [67] (p. 126), including practical examples as well, which has proven to be an effective method. Barth et al. [66] showed a positive effect of a systematic introduction to the relevant knowledge base on both the development of declarative professional knowledge and knowledge-based reasoning skills. Small effects of direct instruction on PCK were also reported by Tröbst et al. [63]. With regard to the development of PCK, the researchers found the combined instruction of the knowledge facets within professional development programs, which considered transformation processes of CK and PK during PCK-construction, to be more effective than polyvalent traditional teacher education [40,63,67]. This combined view on the acquisition of knowledge attributes a high relevance to educational training in higher education. In such trainings, practical classroom scenarios can also be videotaped and used for eliciting reasoning processes about these classroom scenarios, for which positive effects in terms of professional development could be recorded [90,95]. It is therefore important for pre-service teachers to take advantage of corresponding offers of universities.

5. Limitations

First of all, it should be noted that the overall sample size of the reported study was rather small, thus, resulting in an even smaller sample for the five treatments. This is accompanied by losses in terms of statistical power in our calculations. However, since the study was embedded in regular courses within a German university, the number of participants was limited and could not be easily increased up to the time of the study.

Moreover, the increase in professional knowledge in terms of PCK and CK from pre to post might also be the result of the pre/post-test research design, since other experiences made during the study period or other forms of input that took place in everyday life or other seminars may have contributed to teachers' learning (cf. [69]).

Although descriptive PK-measures improved from pre to post, the time effect of the mixed ANOVA did not show statistical significance. One reason might be seen in the subject-specific focus of the videos, in which mainly subject-specific dimensions from biology instruction were addressed. In contrast, studies using videos that focus on general pedagogical aspects showed corresponding effects on PK [26,91]. Another reason for the lack of significant improvement in PK may also be due to the PK test used. Test results for the pre-test were already comparatively high. Thus, the PK test did not discriminate the sample enough. Test reliabilities might be increased by utilizing a longer version of the test, or by a higher number of participants to increase variance in ability [86].

Additionally, the absence of motivational variables that were not considered in this study might be seen as another potentially limiting factor. For a productive measurement, it is important that participants' desire and willingness to solve the tasks are considered and, if necessary, are scaffolded [89]. However, a supplementary analysis considering teachers' situational interest showed that controlling the variable situational interest did not change the results either. Still, an impact on particular diagnostic activities such as *drawing conclusions* may be possible.

A final limitation is of conceptual nature. In our study, we measured content-related knowledge facets (CK, PCK, PK). Theoretically, the different knowledge types *knowing that*, *knowing how*, *knowing when and why* can be distinguished for each of them [10]. However, the utilized PK test merely contained items covering *knowing that*. Additionally, the texts used for the intervention mainly covered declarative knowledge. Even though action-related knowledge in terms of *knowing how* and *knowing when and why* should be prompted by the tasks to be completed in the treatments, their proportion may not have been high enough to significantly influence the diagnostic activities. However, when addressing scientific reasoning, Kind and Osborne [4] emphasized the importance of procedural knowledge comprising the understanding of science-specific procedures (*knowing how*), and epistemic knowledge comprising knowledge about epistemic constructs and values in science (*knowing why*). In future interventions, the proportion of these knowledge types should be expanded accordingly. Moreover, addressing *knowing how* and *knowing when and why* may be better accomplished through other types of instructional support than texts, e.g., through experienced lecturers.

6. Conclusions and Further Research

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

Future research could also investigate additional support in terms of scaffolding within video-based tools or simulations that might further promote the development of knowledge and diagnostic activities in order to facilitate teachers' diagnostic competences [9]. Moreover, scaffolding different types of knowledge relevant for scientific reasoning (that is procedural and epistemic knowledge for problem solving, cf. [4]) may improve scientific reasoning skills such as diagnostic activities and, thus, the development of content-related knowledge as well.

Following the demand for integrated coursework or a combined instruction of the knowledge facets, knowledge acquisition could also be addressed through direct instruction, as it is done, for example, in lectures at universities [66,67]. Therefore, further research could investigate the possible effects of an integrated presentation of the knowledge facets in lectures (cf. [96,97]).

Author Contributions: Conceptualization, M.K., C.F. and B.J.N.; formal analysis, M.K. and C.F.; investigation, M.K.; writing—original draft preparation, M.K.; writing—review and editing, M.K., C.F. and B.J.N.; supervision, C.F. and B.J.N.; project administration, B.J.N..; funding acquisition, B.J.N. All authors have read and agreed to the published version of the manuscript.

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

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

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

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

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

Appendix A

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

Treatment	Content included in the texts	References
1 PCK	- subject-specific features of instructional quality (level of students' cognitive activities, use of originals, use of models, use of experiments and scientific working methods, use of technical language, linking and structuring of content) - lesson planning model for biology instruction considering core ideas	[49,98–112]
2 CK	- basic functions of the skin - structure of the skin - effects of sunlight - skin as a sensory organ - perception of stimuli	[109,113–119]
3 PK	- generic features of instructional quality (basic dimensions: classroom management, supportive climate, cognitive activation) - teaching methods	[109,110,120–125]
4 combination of PCK, CK, PK	- basic dimensions: classroom management, supportive climate, cognitive activation - lesson planning model for biology instruction considering core ideas - basic functions of the skin - structure of the skin - skin as a sensory organ	[99,101–103,105,110–113,115–119]

5 no information (control group)	- reflection on the organisation of the university teacher education and ideas for improvement	-

Appendix B

<u>Tasks</u>:

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

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

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

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

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

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

Mrs. Hanna starts her lesson (on the topic blood groups) by asking the following questions: *Why do you think someone should know their blood group or why should their blood group be known? When might this be important?*
Later she presents the topic of the lesson by saying "And today we want to find out what is meant by the statement *the blood does not match*".
During the rest of the lesson, Mrs. Hanna presents information on different types of blood cells, corresponding antigens and antibodies. The students have to repeat the information afterwards and determine the blood type of six patients.
(Example based on [99].)

b) Describe an alternative teaching strategy in which your criticized aspects are taken into account.

[1] The materials used, including figures and tables, can be made available on request by the authors of this study.

Figure B1. Example of the given tasks utilized in treatment (1) PCK (translated from German).

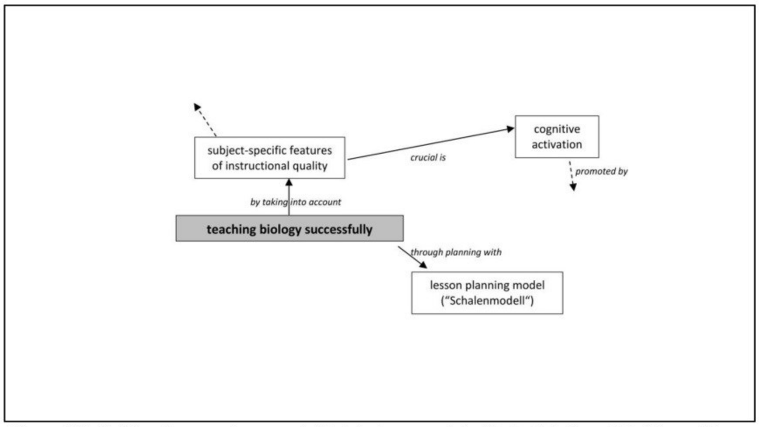

Figure B2. Outlined concept map, which is to be completed in task 2 (translated from German).

References

1. Morris, B.J.; Croker, S.; Masnick, A.M.; Zimmerman, C. The Emergence of Scientific Reasoning. In *Current topics in children's learning and cognition*; Kloos, H., Morris, B.J., Amaral, J.L., Eds.; InTech: Rijeka, Croatia, 2012; pp 61–82, ISBN 978-953-51-0855-9.
2. National Research Council. *Taking science to school: Learning and teaching science in Grades K–8;* National Academies Press: Washington, D.C., 2007.
3. Krell, M.; Redman, C.; Mathesius, S.; Krüger, D.; Van Driel, J. Assessing Pre-Service Science Teachers' Scientific Reasoning Competencies. *Res Sci Educ* **2020**, *50*, 2305–2329, doi:10.1007/s11165-018-9780-1.
4. Kind, P.; Osborne, J. Styles of Scientific Reasoning: A Cultural Rationale for Science Education? *Sci. Ed.* **2017**, *101*, 8–31, doi:10.1002/sce.21251.
5. Mayer, J. Erkenntnisgewinnung als wissenschaftliches Problemlösen [Inquiry as scientific problem solving]. In *Theorien in der biologiedidaktischen Forschung*: *Ein Handbuch für Lehramtsstudenten und Doktoranden [Theories in biology education research. A handbook for pre-service teachers and doctoral students]*; Krüger, D., Vogt, H., Eds.; Springer: Berlin, 2007; pp 177–186.
6. Nowak, K.H.; Nehring, A.; Tiemann, R.; Upmeier zu Belzen, A. Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education* **2013**, *47*, 182–188, doi:10.1080/00219266.2013.822747.
7. Helmke, A.; Lenske, G. Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung [Diagnosis of Classroom Instruction from Different Perspectives as a Prerequisite for Improving Teaching and Learning]. *Beiträge zur Lehrerbildung* **2013**, *31*, 214–233.
8. Kramer, M.; Förtsch, C.; Boone, W.J.; Seidel, T.; Neuhaus, B.J. Investigating Pre-Service Biology Teachers' Diagnostic Competence: Systematic Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy. *Education Sciences* **2021**, *11*, 89, doi:10.3390/educsci11030089.
9. 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 Learning Research* **2019**, *7*, 1–24, doi:10.14786/flr.v7i4.384.

10. Förtsch, C.; Sommerhoff, D.; Fischer, F.; Fischer, M.R.; Girwidz, R.; Obersteiner, A.; Reiss, K.; Stürmer, K.; Siebeck, M.; Schmidmaier, R.; et al. Systematizing Professional Knowledge of Medical Doctors and Teachers: Development of an Interdisciplinary Framework in the Context of Diagnostic Competences. *Education Sciences* **2018**, *8*, 207, doi:10.3390/educsci8040207.
11. Baumert, J.; Kunter, M.; Blum, W.; Brunner, M.; Voss, T.; Jordan, A.; Klusmann, U.; Krauss, S.; Neubrand, M.; Tsai, Y.-M. Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal* **2010**, *47*, 133–180, doi:10.3102/0002831209345157.
12. Tepner, O.; Borowski, A.; Dollny, S.; Fischer, H.E.; Jüttner, M.; Kirschner, S.; Leutner, D.; Neuhaus, B.J.; Sandmann, A.; Sumfleth, E.; et al. Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften [Item Development Model for Assessing Professional Knowledge of Science Teachers]. *ZfDN* **2012**, *18*.
13. Ball, D.L.; Thames, M.H.; Phelps, G. Content Knowledge for Teaching: What Makes It Special? *Journal of Teacher Education* **2008**, *59*, 389–407, doi:10.1177/0022487108324554.
14. *TEDS-M 2008. Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich [TEDS-M 2008. Professional competence and learning opportunities of prospective secondary mathematics teachers in international comparison];* Blömeke, S.; Kaiser, G.; Lehmann, R., Eds.; Waxmann: Münster, 2010.
15. Blömeke, S.; Busse, A.; Kaiser, G.; König, J.; Suhl, U. The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education* **2016**, *56*, 35–46, doi:10.1016/j.tate.2016.02.003.
16. Hoth, J.; Döhrmann, M.; Kaiser, G.; Busse, A.; König, J.; Blömeke, S. Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM Mathematics Education* **2016**, *48*, 41–53, doi:10.1007/s11858-016-0759-y.
17. 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. *American Educational Research Journal* **2012**, *49*, 568–589.
18. Blömeke, S.; Gustafsson, J.-E.; Shavelson, R.J. Beyond Dichotomies. *Zeitschrift für Psychologie* **2015**, *223*, 3–13, doi:10.1027/2151-2604/a000194.
19. Boyatzis, R.E. *The competent manager;* Wiley: New York, NY, 1982.
20. 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, doi:10.1007/s10649-016-9724-5.
21. Depaepe, F.; Verschaffel, L.; Kelchtermans, G. Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education* **2013**, *34*, 12–25, doi:10.1016/j.tate.2013.03.001.
22. Zimmerman, C. The Development of Scientific Reasoning Skills. *Developmental Review* **2000**, *20*, 99–149, doi:10.1006/drev.1999.0497.
23. Dorfner, T.; Förtsch, C.; Boone, W.J.; Neuhaus, B.J. Instructional Quality Features in Videotaped Biology Lessons: Content-Independent Description of Characteristics. *Res Sci Educ* **2019**, *49*, 1457–1491, doi:10.1007/s11165-017-9663-x.
24. Lipowsky, F.; Rakoczy, K.; Pauli, C.; Drollinger-Vetter, B.; Klieme, E.; Reusser, K. Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction* **2009**, *19*, 527–537, doi:10.1016/j.learninstruc.2008.11.001.

25. Seidel, T.; Shavelson, R.J. Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research* **2007**, *77*, 454–499, doi:10.3102/0034654307310317.
26. König, J.; Kramer, C. Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM Mathematics Education* **2016**, *48*, 139–151, doi:10.1007/s11858-015-0705-4.
27. 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, 2016; pp 77–94, ISBN 9781315749136.
28. Baumert, J.; Kunter, M. The COACTIV Model of Teachers' Professional Competence. In *Cognitive activation in the mathematics classroom and professional competence of teachers*: *Results from the COACTIV project*; Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., Neubrand, M., Eds.; Springer: New York, 2013; pp 25–48, ISBN 978-1-4614-5148-8.
29. Weinert, F.E.; Schrader, F.-W.; Helmke, A. Educational expertise. Closing the gap between educational research and classroom practice. *School Psychology International* **1990**, *11*, 163–180.
30. 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 Learning Research* **2014**, *5*, 28–45, doi:10.14786/flr.v2i3.96.
31. Paris, S.G.; Lipson, M.Y.; Wixson, K.K. Becoming a Strategic Reader. *Contemporary Educational Psychology* **1983**, *8*, 293–316.
32. Shulman, L.S. Those who understand: Knowledge growth in teaching. *Educational Researcher* **1986**, *15*, 4–14.
33. Shulman, L.S. Knowledge and teaching: Foundations of the new reform. *Havard Educational Review* **1987**, *57*, 1–22.
34. 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 Netherlands: Dordrecht, 2012; pp 435–448, ISBN 978-1-4020-9040-0.
35. Voss, T.; Kunter, M.; Baumert, J. Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology* **2011**, *103*, 952–969, doi:10.1037/a0025125.
36. 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, doi:10.1026/0012-1924/a000238.
37. 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, 2015; pp 28–42.
38. 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'. *International Journal of Science and Mathematics Education* **2013**, *11*, 1369–1390.
39. Alonzo, A.C.; Berry, A.; Nilsson, P. Unpacking the Complexity of Science Teachers' PCK in Action: Enacted and Personal PCK. In *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*; Hume, A., Cooper, R., Borowski, A., Eds.; Springer: Singapore, 2019; pp 271–286.

40. Gess-Newsome, J.; Taylor, J.A.; Carlson, J.; Gardner, A.L.; Wilson, C.D.; Stuhlsatz, M.A.M. Teacher pedagogical content knowledge, practice, and student achievement. *International Journal of Science Education* **2019**, *41*, 944–963, doi:10.1080/09500693.2016.1265158.
41. Carlson, J.; Daehler, K.R.; Alonzo, A.C.; Barendsen, E.; Berry, A.; Borowski, A.; Carpendale, J.; Kam Ho Chan, K.; Cooper, R.; Friedrichsen, P.; et al. The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*; Hume, A., Cooper, R., Borowski, A., Eds.; Springer: Singapore, 2019; pp 77–94.
42. Magnusson, S.; Krajcik, J.; Borko, H. Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching. In *Examining Pedagogical Content Knowledge: The Construct and its Implications for Science Education*; Gess-Newsome, J., Lederman, N.G., Eds.; Springer: Dordrecht, 1999; pp 95–132.
43. Jüttner, M.; Boone, W.J.; Park, S.; Neuhaus, B.J. Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educ Asse Eval Acc* **2013**, *25*, 45–67, doi:10.1007/s11092-013-9157-y.
44. Shavelson, R.J. On the measurement of competency. *Empirical research in vocational education and training* **2010**, *2*, 41–63.
45. Brown, N.J.S.; Furtak, E.M.; Timms, M.; Nagashima, S.O.; Wilson, M. The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning. *Educational Assessment* **2010**, *15*, 123–141, doi:10.1080/10627197.2010.530551.
46. Dorfner, T.; Förtsch, C.; Germ, M.; Neuhaus, B.J. Biology instruction using a generic framework of scientific reasoning and argumentation. *Teaching and Teacher Education* **2018**, *75*, 232–243, doi:10.1016/j.tate.2018.07.003.
47. European Commission. *Science education for responsible citizenship: Report to the European Commission of the Expert Group on Science Education;* Publications Office of the European Union: Luxembourg, 2015, ISBN 9789279436369.
48. Wildgans-Lang, A.; Scheuerer, S.; Obersteiner, A.; Fischer, F.; Reiss, K. Analyzing prospective mathematics teachers' diagnostic processes in a simulated environment. *ZDM Mathematics Education* **2020**, *57*, 175, doi:10.1007/s11858-020-01139-9.
49. Förtsch, C.; Werner, S.; Kotzebue, L. von; Neuhaus, B.J. Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education* **2016**, *38*, 2642–2666, doi:10.1080/09500693.2016.1257170.
50. Loibl, K.; Leuders, T.; Dörfler, T. A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teaching and Teacher Education* **2020**, *91*, 103059, doi:10.1016/j.tate.2020.103059.
51. Ohle, A.; McElvany, N.; Horz, H.; Ullrich, M. Text-picture integration - Teachers' attitudes, motivation and self-related cognitions in diagnostics. *Journal for Educational Research Online* **2015**, *7*, 11–33.
52. Ostermann, A.; Leuders, T.; Nückles, M. Improving the judgment of task difficulties: prospective teachers' diagnostic competence in the area of functions and graphs. *J Math Teacher Educ* **2018**, *21*, 579–605, doi:10.1007/s10857-017-9369-z.
53. Kim, S.; Rehder, B. How prior knowledge affects selective attention during category learning: an eyetracking study. *Mem. Cognit.* **2011**, *39*, 649–665, doi:10.3758/s13421-010-0050-3.
54. Bauer, E.; Fischer, F.; Kiesewetter, J.; Shaffer, D.W.; Fischer, M.R.; Zottmann, J.M.; Sailer, M. Diagnostic Activities and Diagnostic Practices in Medical Education and Teacher Education. *Front. Psychol.* **2020**.
55. Kramer, M.; Förtsch, C.; Seidel, T.; Neuhaus, B.J. Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation* **2021**, *28*, doi:10.1016/j.stueduc.2020.100973.

56. 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 of Sci and Math Educ* **2015**, *13*, 369–387, doi:10.1007/s10763-015-9616-7.
57. Seidel, T.; Stürmer, K. Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal* **2014**, *51*, 739–771, doi:10.3102/0002831214531321.
58. Sherin, M.G.; van Es, E.A. Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education* **2009**, *60*, 20–37, doi:10.1177/0022487108328155.
59. Hetmanek, A.; Engelmann, K.; Opitz, A.; Fischer, F. *Beyond intelligence and domain knowledge: Scientific reasoning and argumentation as a set of cross-domain skills. Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge;* Routledge: New York, 2018.
60. Klemm, J.; Flores, P.; Sodian, B.; Neuhaus, B.J. Scientific Reasoning in Biology – the Impact of Domain-General and Domain-Specific Concepts on Children's Observation Competency. *Front. Psychol.* **2020**, *11*, 53, doi:10.3389/fpsyg.2020.01050.
61. Krell, M.; Reinisch, B.; Krüger, D. Analyzing Students' Understanding of Models and Modeling Referring to the Disciplines Biology, Chemistry, and Physics. *Res Sci Educ* **2015**, *45*, 367–393, doi:10.1007/s11165-014-9427-9.
62. 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. *Teaching and Teacher Education* **2017**, *66*, 158–170, doi:10.1016/j.tate.2017.04.010.
63. Tröbst, S.; Kleickmann, T.; Depaepe, F.; Heinze, A.; Kunter, M. Effects of instruction on pedagogical content knowledge about fractions in sixth-grade mathematics on content knowledge and pedagogical knowledge. *Unterrichtswiss* **2019**, *47*, 79–97, doi:10.1007/s42010-019-00041-y.
64. Gruber, H.; Mandl, H.; Renkl, A. Was lernen wir in der Schule und Hochschule: Träges Wissen? [What do we learn in school and college: inert knowledge?]. In *Die Kluft zwischen Wissen und Handeln. Empirische und theoretische Lösungsansätze [The gap between knowledge and action. Empirical and Theoretical Approaches]*; Mandl, H., Gerstenmaier, J., Eds.; Hogrefe: Göttingen, 2000; 139-156.
65. Meier, M.; Grospietsch, F.; Mayer, J. Vernetzung von Wissensfacetten professioneller Handlungskompetenz in hochshuldidaktischen Lehr-Lernsettings [Linking of knowledge facets of professional acting in teaching-learning settings of higher education]. In *Kohärenz in der universitären Lehrerbildung*; Glowinski, I., Borowski, A., Gillen, J., Schanze, S., Meien, J. von, Eds.; Universitätsverlag Potsdam: Potsdam, 2018; pp 143–178.
66. 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. *International Journal of Educational Research* **2019**, *95*, 1–12, doi:10.1016/j.ijer.2019.03.002.
67. Kleickmann, T.; Tröbst, S.; Heinze, A.; Bernholt, A.; Rink, R.; Kunter, M. Teacher Knowledge Experiment: Conditions of the Development of Pedagogical Content Knowledge. In *Competence assessment in education: Research, models and instruments*; Leutner, D., Fleischer, J., Grünkorn, J., Klieme, E., Eds.; Springer: Cham, 2017; pp 111–129, ISBN 978-3-319-50028-7.
68. Smit, R.; Weitzel, H.; Blank, R.; Rietz, F.; Tardent, J.; Robin, N. Interplay of secondary pre-service teacher content knowledge (CK), pedagogical content knowledge (PCK) and attitudes regarding scientific inquiry teaching within teacher training. *Research in Science & Technological Education* **2017**, *35*, 477–499, doi:10.1080/02635143.2017.1353962.
69. Santagata, R.; Guarino, J. Using video to teach future teachers to learn from teaching. *ZDM Mathematics Education* **2011**, *43*, 133–145, doi:10.1007/s11858-010-0292-3.

70. van Es, E.A.; Sherin, M.G. Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education* **2008**, *24*, 244–276, doi:10.1016/j.tate.2006.11.005.
71. Furtak, E.M.; Bakeman, R.; Buell, J.Y. Developing knowledge-in-action with a learning progression: Sequential analysis of teachers' questions and responses to student ideas. *Teaching and Teacher Education* **2018**, *76*, 267–282, doi:10.1016/j.tate.2018.06.001.
72. Kramer, M.; Förtsch, C.; Stürmer, J.; Förtsch, S.; Seidel, T.; Neuhaus, B.J. Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education* **2020**, *7*, doi:10.1080/2331186X.2020.1823155.
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, 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, doi:10.1111/j.1440-1754.2011.02115.x.
75. Jüttner, M.; Neuhaus, B.J. Development of Items for a Pedagogical Content Knowledge Test Based on Empirical Analysis of Pupils' Errors. *International Journal of Science Education* **2012**, *34*, 1125–1143, doi:10.1080/09500693.2011.606511.
76. 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, 2002, ISBN 3801716465.
77. Kunter, M.; Leutner, D.; Terhart, E.; Baumert, J. Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung (BilWiss) [Broad Pedagogical Knowledge and the Development of Professional Competence in Teacher Education (BilWiss)] (Version 5) [Data set]. Available online: https://www.iqb.hu-berlin.de/fdz/studies/BilWiss (accessed on 10 February 2021).
78. 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, 2001; pp 43–57.
79. Bond, T.G.; Fox, C.M.F. *Applying the Rasch model: Fundamental measurement in the human sciences*, 2nd ed.; Lawrence Erlbaum: Mahwah, New Jersey, 2007.
80. Boone, W.J.; Staver, J.; Yale, M. *Rasch analysis in the human sciences;* Springer: Dordrecht, 2014, ISBN 9789400768574.
81. Wright, B.D.; Linacre, J.M. Reasonable mean-square fit values. *Rasch Measurement Transactions* **1994**, *8*.
82. 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).
83. Questback GmbH. *EFS Survey;* Questback GmbH: Köln, 2018.
84. 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, doi:10.1007/s40573-017-0058-3.
85. Kersting, N.B.; Sherin, B.L.; Stigler, J.W. Automated Scoring of Teachers' Open-Ended Responses to Video Prompts. *Educational and Psychological Measurement* **2014**, *74*, 950–974, doi:10.1177/0013164414521634.

86. Linacre, J.M. A User's Guide to Winsteps® Ministeps Rasch-Model Computer Programs: Program Manual 3.81.0. Available online: http://www.winsteps.com/manuals.htm.
87. Kyriakides, L.; Christoforou, C.; Charalambous, C.Y. What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education* **2013**, *36*, 143–152, doi:10.1016/j.tate.2013.07.010.
88. 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. *Teaching and Teacher Education* **2014**, *38*, 76–88, doi:10.1016/j.tate.2013.11.004.
89. Belland, B.R.; Kim, C.; Hannafin, M.J. A Framework for Designing Scaffolds That Improve Motivation and Cognition. *Educational Psychologist* **2013**, *48*, 243–270, doi:10.1080/00461520.2013.838920.
90. Osborne, J.F.; Borko, H.; Fishman, E.; Gomez Zaccarelli, F.; Berson, E.; Busch, K.C.; Reigh, E.; Tseng, A. Impacts of a Practice-Based Professional Development Program on Elementary Teachers' Facilitation of and Student Engagement With Scientific Argumentation. *American Educational Research Journal* **2019**, *56*, 1067–1112, doi:10.3102/0002831218812059.
91. Stürmer, K.; Könings, K.D.; Seidel, T. Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *Br. J. Educ. Psychol.* **2012**, *83*, 467–483, doi:10.1111/j.2044-8279.2012.02075.x.
92. van Es, E.A.; Sherin, M.G. Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher* **2002**, *10*, 571–596.
93. Santagata, R.; Yeh, C. The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM Mathematics Education* **2016**, *48*, 153–165, doi:10.1007/s11858-015-0737-9.
94. Kaiser, G.; Blömeke, S.; Busse, A.; Döhrmann, M.; König, J. Professional knowledge of (prospective) Mathematics teachers – Its structure and development. In *Proceedings of the joint meeting of PME 38 and PME-NA 36*; Liljedahl, P., Nicol, C., Oesterle, S., Allan, Eds.; PME: Vancouver, 2014; pp 35–50.
95. 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]. *ZfDN* **2019**, *42*, 271, doi:10.1007/s40573-019-00103-9.
96. Harr, N.; Eichler, A.; Renkl, A. Integrating pedagogical content knowledge and pedagogical/psychological knowledge in mathematics. *Front. Psychol.* **2014**, *5*, 1–10, doi:10.3389/fpsyg.2014.00924.
97. Harr, N.; Eichler, A.; Renkl, A. Integrated learning: ways of fostering the applicability of teachers' pedagogical and psychological knowledge. *Front. Psychol.* **2015**, *6*, 1–16, doi:10.3389/fpsyg.2015.00738.
98. Neuhaus, B.J.; Nachreiner, K.; Oberbei, I.; Spangler, M. Basiskonzepte zur Planung von Biologieunterricht. Ein Gedankenspiel [Core ideas for biology lesson planning. A mental game]. *MNU (Mathematisch-naturwissenschaftlicher Unterricht)* **2014**, *67*, 160–163.
99. Wadouh, J.; Liu, N.; Sandmann, A.; Neuhaus, B.J. The Effect of Knowledge linking Levels in Biology Lessons upon Students' Knowledge Structure. *International Journal of Science and Mathematics Education* **2014**, *12*, 25–47, doi:10.1007/s10763-012-9390-8.
100. Förtsch, C.; Werner, S.; Dorfner, T.; Kotzebue, L. von; Neuhaus, B.J. Effects of Cognitive Activation in Biology Lessons on Students' Situational Interest and Achievement. *Res Sci Educ* **2017**, *47*, 559–578, doi:10.1007/s11165-016-9517-y.

101. Nawani, J.; Kotzebue, L.; Rixius, J.; Graml, M.; Neuhaus, B.J. Teachers' Use of Focus Questions in German Biology Classrooms: a Video-based Naturalistic Study. *Int J of Sci and Math Educ* **2017**, *95*, 639, doi:10.1007/s10763-017-9837-z.
102. Dorfner, T.; Förtsch, C.; Neuhaus, B.J. Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction. *Learning and Instruction* **2018**, *56*, 42–53, doi:10.1016/j.learninstruc.2018.03.001.
103. Dorfner, T.; Förtsch, C.; Neuhaus, B.J. Use of Technical Terms in German Biology Lessons and its Effects on Students' Conceptual Learning. *Research in Science & Technological Education* **2019**, *38*, 227–251, doi:10.1080/02635143.2019.1609436.
104. Förtsch, C.; Dorfner, T.; Baumgartner, J.; Werner, S.; Kotzebue, L. von; Neuhaus, B.J. Fostering Students' Conceptual Knowledge in Biology in the Context of German National Education Standards. *Res Sci Educ* **2020**, *50*, 739–771, doi:10.1007/s11165-018-9709-8.
105. Haugwitz, M. Kontextorientiertes Lernen und Concept Mapping im Fach Biologie. Eine experimentelle Untersuchung zum Einfluss auf Interesse und Leistung unter Berücksichtigung von Moderationseffekten individueller Voraussetzungen beim kooperativen Lernen [Context-oriented learning and concept mapping in biology. An experimental investigation of the influence on interest and achievement considering moderation effects of individual prerequisites in cooperative learning]. Dissertation; Universität Duisburg-Essen, Duisburg, 2009.
106. Jatzwauk, P.; Rumann, S.; Sandmann, A. Der Einfluss des Aufgabeneinsatzes im Biologieunterricht auf die Lernleistung der Schüler – Ergebnisse einer Videostudie [The effect of usage of tasks in biology education on learning performance – a video study]. *ZfDN* **2008**, *14*, 263-282.
107. Kattmann, U. Modelle. In *Fachdidaktik Biologie,* 8. Auflage; Gropengießer, H., Kattmann, U., Eds.; Aulis: Köln, 2008; 330-339.
108. Klieme, E.; Pauli, C.; Reusser, K. The Pythagoras Study: Investigating Effects of Teaching and Learning in Swiss and German Mathematics Classrooms. In *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*; Janík, T., Seidel, T., Eds.; Waxmann: Münster u.a., 2009; pp 137–160.
109. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Standards für die Lehrerbildung - Bildungswissenschaften. Available online: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf (accessed on 16 October 2020).
110. Seidel, T.; Schindler, A.-K. Klassenführung. In *Handwörterbuch Pädagogische Psychologie [Handbook of Educational Psychology]*, 5th, revised and expanded edition; Rost, D.H., Sparfeldt, J.R., Buch, S.R., Eds.; Beltz: Weinheim, 2018; pp 328–336, ISBN 978-3-621-28297-0.
111. Werner, S. Zusammenhänge zwischen dem fachspezifischen Professionswissen einer Lehrkraft, dessen Unterrichtsgestaltung und Schülervariablen am Beispiel eines elaborierten Modelleinsatzes. Dissertation; Ludwig-Maximilians Universität, München, 2016.
112. Wüsten, S. Allgemeine und fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie: Eine Video- und Interventionsstudie [General and subject-specific instructional quality in the subject biology. A video and intervention study]. Dissertation; Universität Duisburg-Essen, 2010.
113. Hickman, C.P.; Roberts, L.S.; Larson, A.; L'Anson, H.; Eisenhour, D.J. *Zoologie [Zoology]*, 13th, updated ed.; Pearson Education: München, 2008, ISBN 3827372658.

114. Hildebrand, M.; Goslow, G.E. *Vergleichende und funktionelle Anatomie der Wirbeltiere [Comparative and functional anatomy of vertebrates]*; Springer: Berlin, Heidelberg, 2004, ISBN 9783642189517.
115. Mörike, K.D.; Betz, E.; Mergenthaler, W. *Biologie des Menschen [human biology]*, 15th ed.; Nikol: Hamburg, 2007, ISBN 9783937872551.
116. Moyes, C.D.; Schulte, P.M. *Tierphysiologie [Animal Physiology]*, 1. Aufl.; Pearson Studium: München, Boston [u.a.], 2008, ISBN 9783827372703.
117. Müller, W.; Frings, S. *Tier- und Humanphysiologie. Eine Einführung [Animal and human physiology. An introduction]*, 3rd, revised and updated edition; Springer: Berlin, Heidelberg, 2007, ISBN 9783540327332.
118. Purves, W.K.; Sadava, D.; Orians, G.H.; Heller, H.C. *Biologie [Biology]*, 7th ed.; Elsevier: Munich, 2004.
119. Thews, G.; Mutschler, E.; Vaupel, P. *Anatomie Physiologie Pathophysiologie des Menschen. Ein Lehrbuch für Pharmazeuten und Biologen [Anatomy Physiology Pathophysiology of the human being. A textbook for pharmacists and biologists]*; Wissenschaftliche Verlagsgesellschaft mbH.: Stuttgart, 1980.
120. Eder, F. Schul- und Klassenklima. In *Handwörterbuch Pädagogische Psychologie [Handbook of Educational Psychology]*, 5th, revised and expanded edition; Rost, D.H., Sparfeldt, J.R., Buch, S.R., Eds.; Beltz: Weinheim, 2018; pp 696–707, ISBN 978-3-621-28297-0.
121. Helmke, A. Aktive Lernzeit optimieren – Was wissen wir über effiziente Klassenführung? [Optimizing Time on Task - What do we know about effective classroom management?]. *Pädagogik* **2007**, *59*, 44–49.
122. Helmke, A. *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts [Quality of Teaching and Teacher Professionalism. Diagnosis, Evaluation and Improvement of teaching]*, 5th ed.; Klett-Kallmeyer: Seelze, 2014.
123. Helmke, A.; Brühwiler, C. Unterrichtsqualität [Instructional Quality]. In *Handwörterbuch Pädagogische Psychologie [Handbook of Educational Psychology]*, 5th, revised and expanded edition; Rost, D.H., Sparfeldt, J.R., Buch, S.R., Eds.; Beltz: Weinheim, 2018; pp 860–869, ISBN 978-3-621-28297-0.
124. *Biologieunterricht heute. Eine moderne Fachdidaktik [Biology classes today. A modern view on biology education]*; Killermann, W.; Hiering, P.; Starosta, B., Eds., 12th ed.; Auer: Donauwörth, 2008.
125. Koran, M.L.; Koran, J.J. Aptitude-treatment interaction research in science education. *Journal of Research in Science Teaching* **1984**, *21*, 793–808, doi:10.1002/tea.3660210804.

4. Discussion

In the following, the results of all publications and manuscripts are first summarized before they are discussed with regard to the aims of this dissertation. The discussion starts with the validation of the video-based assessment tool DiKoBi *Assess* and the comparison of the constructs *professional vision* and *diagnostic activities* as feasible operationalizations of situation-specific skills required in (video-based) problem-solving situations. Then, the relationships that were found between the components of diagnostic competences are described and discussed in terms of previous research findings. In addition, the effectiveness of instructional support provided by a lecturer or via texts, as well as their impact on the professional knowledge base as an important component of diagnostic competences is discussed. The discussion of the results is followed by elaborating on the limitations of the studies, which lead to the description of further research questions. Finally, implications for teacher education that can be derived from the results of this dissertation are considered.

4.1 Summary of the results

This dissertation pursued three aims that were addressed in three publications and two manuscripts: (1) to develop and validate a video-based tool to assess pre-service biology teachers' diagnostic competences, (2) to investigate the relationships between professional knowledge, diagnostic activities, and diagnostic accuracy as components of diagnostic competences, and (3) to analyze the effects of different types of instructional support on the professional knowledge base representing a crucial component of diagnostic competences. Figure 6 presents the main results of this dissertation:

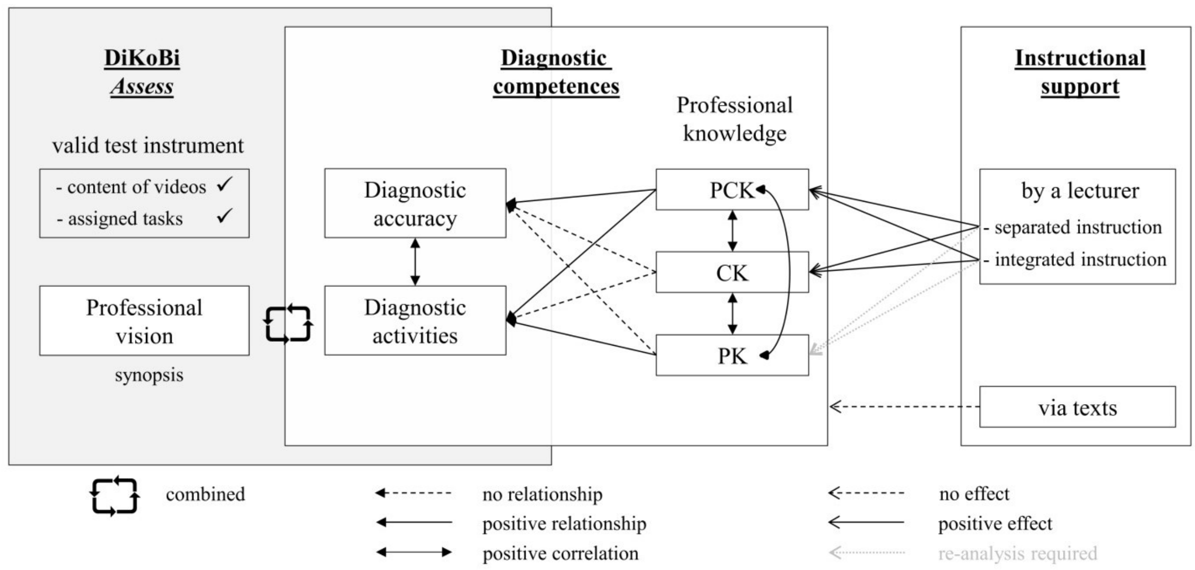

Figure 6. Summary of the main results of this dissertation.

The validation of the video-based tool DiKoBi *Assess* with in-service biology teachers showed that DiKoBi *Assess* could be used as a valid test instrument. This is due to the fact that the staged videos were able to simulate biology-specific classroom situations authentically. Moreover, after conducting several validation steps, it could be concluded that teachers sufficiently identified the scripted subject-specific characteristics (*content of videos*) in each classroom situation and that the *assigned tasks* validly measured the assumed diagnostic activities (see Appendix A for an example of the tasks embedded in DiKoBi *Assess*). Additionally, the comparison of the constructs *professional vision* and *diagnostic activities*, which were both used independently for coding and analyzing the data collected in DiKoBi *Assess*, showed that with regard to the diagnostic activities relevant in the context of the tool, both constructs provide comparable coding and analysis results. Further details of the comparison are described in Section 4.2.

Based on the validation, DiKoBi *Assess* was used in two studies as a measurement instrument for diagnostic activities and diagnostic accuracy in addition to conventional paper-pencil tests measuring professional knowledge. The analyzed relationships between the three components of diagnostic competences revealed a strong correlation between diagnostic activities and diagnostic accuracy, a moderate correlation between the knowledge facets PCK and CK, and small correlations between PCK and PK as well as between PK and CK. In addition, there was a positive directional relationship between PCK and pre-service teachers' diagnostic activities, PCK and diagnostic accuracy, and PK and diagnostic activities. Overall, the results highlighted PCK and PK as particularly important for performing diagnostic activities and accurate diagnoses, with higher effect sizes for PCK compared to those of PK.

In terms of instructional support, both separated and integrated instruction provided by a lecturer has been effective in fostering pre-service teachers' PCK and CK. However, only for CK, there were large, statistically significant interaction effects between time and treatments for both the separated and the integrated instruction compared to the control group that showed no significant increase in CK. The descriptive results indicated slightly higher effectiveness of separated instruction for CK development. For PCK, there was a statistically significant increase for all treatments (integrated instruction, separated instruction, and no instruction). However, it is particularly noteworthy that a significant increase in PCK also occurred in the control group, which received no instruction. Descriptive results indicated that addressing the knowledge facets in an interrelated way might be beneficial for the development of PCK. However, this assumption needs to be further examined. In contrast, no effects of instructional support by a lecturer were found for PK, possibly due to a ceiling effect caused by the knowledge test used.

Instructional support via texts did not seem to be effective for any of the components of diagnostic competences within this particular study design since an interaction effect between time and treatment was not found for any of the treatments. At the same time, however, there were significant time effects visible in the increases in PCK and CK. Therefore, it becomes apparent that in both the lecture-based and the text-based intervention study, there must be another factor that led to the knowledge increases in the subject-specific knowledge facets PCK and CK. The assumption is that these increases may be due to the processing of the videos in the video-based assessment tool DiKoBi *Assess* showing biology-specific classroom situations. This assumption is discussed in Section 4.4 in more detail.

4.2 Validation of the video-based assessment tool DiKoBi *Assess*

The first aim of this dissertation was the validation of the video-based assessment tool DiKoBi *Assess* that provided subject-specific classroom situations. The videos embedded in DiKoBi *Assess* were designed to show biology-specific instructional challenges of a biology teacher's classroom behavior. The diagnostic task required the diagnosis of this behavior with regard to subject-specific features of instructional quality. One novelty of this approach was that the diagnostic focus was on the whole class and the instructional behavior of the teacher, which is considered more complicated than diagnosing an individual student (Tolsdorf & Markic, 2017). On the other hand, the tool allowed the identification and interpretation of subject-specific dimensions of instructional quality through practice-oriented cases. Previous tools for assessing diagnostic skills or diagnostic accuracy rather focused on the diagnosis of student performance, on teacher-student interaction, or general characteristics such as classroom management (e.g., Codreanu et al., 2020; Fischer et al., 2021; Gold & Holodynski, 2017; Südkamp et al., 2008). Previous research in biology education has been conducted in the context of the simulated classroom biology SCR^{Bio} (cf. Fischer et al., 2018) and *the student inventory* for biology (cf. Fischer et al., 2021), in which diagnostic competences are investigated with respect to declarative and action-related PK, CK, and PCK. However, the diagnostic focus in these studies was also on measures of student performance and products. Skills such as professional vision or diagnostic activities have rarely been conceptualized and studied in relation to aspects of instructional quality in subject-specific domains (Meschede et al., 2017). Therefore, being able to validly measure situation-specific skills in terms of diagnostic activities with DiKoBi *Assess* extends existing measurement instruments not only by the range of skills that can be measured with these instruments but also by the biology context to which the skills are applied.

In addition, DiKoBi *Assess* is an important instrument for conducting studies within the COSIMA project (cf. Heitzmann et al., 2019; Chernikova et al., in press). Furthermore, the validation in terms of the utilized construct of *diagnostic activities* used for coding and analyzing the data as well as the conceptual comparison with *professional vision* can be considered meaningful for the connection of the interdisciplinary research project to teacher professionalism research and teacher education. In video-based teacher research, *professional vision* is a common concept used to examine teachers' situation-specific skills for classroom assessment and teachers' situated professional knowledge to promote them in terms of teachers' professional development (e.g., Kersting et al., 2012; Seidel & Stürmer, 2014; Sherin & van Es, 2009; Steffensky et al., 2015). The comparison of the constructs *diagnostic activities* and *professional vision* confirmed on the one hand that both constructs are basically suitable for coding situation-specific skills of teachers and that utilizing the constructs for analyses leads to comparable conclusions (Kramer, Förtsch, Seidel et al., 2021). On the other hand, it also became apparent that specific particularities were present in a small percentage of the data, which could be better captured by one of the constructs, depending on the particularity. Adjusting to our project-specific data, relevant categories from both constructs were therefore combined by merging similar construct categories from both constructs as well as subdividing categories more finely with respect to the specific particularities. The resulting combined conceptualization (see Figure 7) was then used as the conceptual basis for further data analysis in the subsequent COSIMA studies. With regard to the tasks embedded in DiKoBi *Assess* that prompted specific diagnostic activities, three categories accounted for the majority of coded responses. These categories were “describing as generating evidence”, “explaining as linking evidence to theory“, and “concluding alternative teaching strategies”. In terms of the construct *diagnostic activities*, the categories represent operationalizations of the diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions*. These main categories are also consistent with subdivisions of other studies, such as those used in the PID model (Kaiser et al., 2015). Furthermore, it is worth mentioning that only a few studies have used videos to promote teachers’ decision-making so far (cf. Santagata et al., 2021), so that addressing the skill “concluding alternative teaching strategies” in the video-based tool DiKoBi *Assess* will provide empirical data for a more detailed analysis of this skill in future studies.

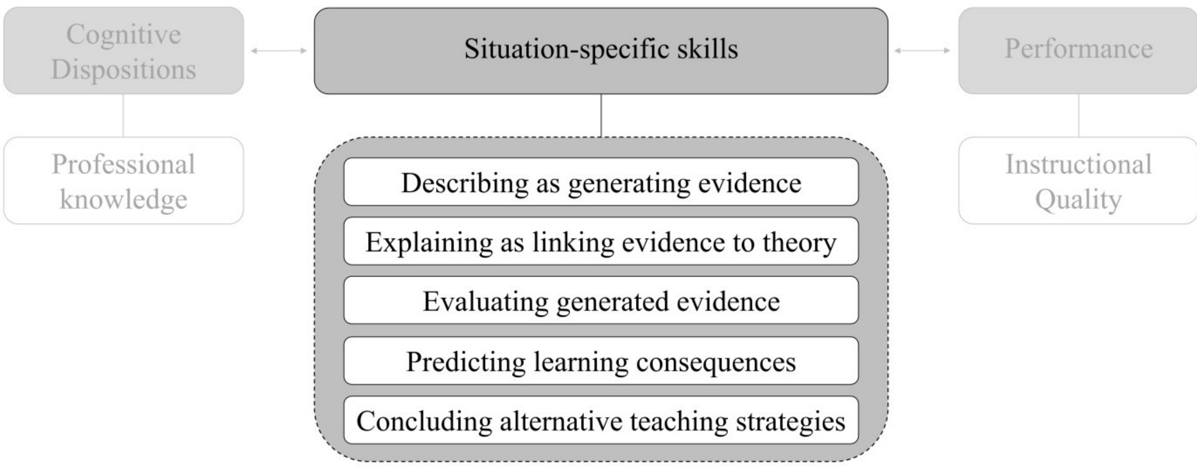

Figure 7. Combined conceptualization for situation-specific skills for diagnosing classroom situations from Kramer, Förtsch, Seidel et al. (2021), embedded in the competence as a continuum model (Blömeke et al., 2015).

4.3 Relationships between components of diagnostic competences

The results of the cross-sectional analysis provided answers regarding the second aim of the dissertation. First, the analysis of the pre-tests from Studies 1 and 2 revealed a just moderate correlation between PCK and CK ($r = 0.30$), whereas results from other studies in mathematics and science education showed moderate to high correlations (e.g., Großschedl, Mahler et al., 2014; Kirschner et al., 2016; Kleickmann et al., 2017; Kulgemeyer et al., 2020; Sorge, Keller et al., 2019). The just moderate correlation from the cross-sectional analysis can be attributed to the underlying characteristics of test construction. Similar knowledge tests have been used in the ProwiN project. The aim of the ProwiN project was to find predictors of instructional quality. Therefore, predictors that do not correlate were better suited for testing. This was artificially achieved by the way the test was constructed, for example, by providing the CK content in the PCK test in order to be able to answer PCK questions as independently of CK as possible (Jüttner & Neuhaus, 2012). Studies in which the ProwiN knowledge tests were utilized confirmed a largely independent measurement of CK and PCK (cf. Förtsch et al., 2016; Jüttner et al., 2013). The moderate correlation found in the cross-sectional analysis might be due to the test versions adapted in the COSIMA project, which concerned different subject content.

Although CK is not directly related to diagnostic activities or accuracy, researchers in mathematics and science education emphasized its role for PCK development (Fischer et al., 2012; Krauss et al., 2008). Furthermore, there was a correlation between PCK and PK, only slightly lower than the one between PCK and CK. This correlation can be attributed to an overlap of PCK and PK test items (see next paragraphs). In addition, there is evidence that for pre-service teachers, PCK and PK initially correlate higher, and as expertise increases, the correlation turns in favor of PCK and CK (Sorge, Stender et al., 2019). This points to the importance of PK in the early stages of pre-service teacher education (cf. König et al., 2011). However, such a shift in correlation during the course of teacher education is probably not measurable with the test instruments used in this dissertation due to the underlying characteristics of test construction.

Regarding PK and CK, a very small correlation was found. Depending on the test construction, similar results on the correlations between PCK/CK and PK were also found in other studies (e.g., Kirschner et al., 2016; Kleickmann et al., 2017; Kulgemeyer et al., 2020).

The strong correlation between diagnostic activities and accuracy might be due to the operationalization of accuracy that was measured in terms of perception accuracy, related to the correct perception and description of a professional situation (cf. Carter et al., 1988; Kaiser et al., 2015). The assessment of accuracy was thus primarily related to the diagnostic activities *identifying problems* and *generating evidence*, which were used to decide whether the scripted biology-specific instructional challenges were correctly diagnosed.

Nevertheless, for both diagnostic activities and accuracy, pre-service teachers' PCK was significantly important. Thus, the results support the tendency in previous findings emphasizing the relation between PCK and situation-specific skills in domain-specific situations (e.g., Dreher & Kuntze, 2015; Kersting et al., 2012; Meschede et al., 2017). Furthermore, the positive relationship between PCK and diagnostic accuracy is important to consider. This result is in line with the expectations, as the diagnostic focus and thus the study context was on subject-specific dimensions of instructional quality in biology, thus requiring the activation of content-specific knowledge. However, this also means that depending on the diagnostic focus, a different knowledge facet may be crucial for the interplay between the components of diagnostic competences (cf. Hoth et al., 2018; Steffensky et al., 2015). For example, teachers' PK has been found critical for interpreting general pedagogical classroom situations (König et al., 2014). However, in the real classroom, content-specific and general pedagogical aspects are not always strictly separable but build on each other to a certain extent, as general pedagogical principles and strategies are sharpened for application in subject-specific situations with regard to the specific content. This intertwining becomes clear, for example, with regard to the basic dimension *cognitive activation*, which is considered to be a generic feature of instructional quality, but has to be refined with regard to the specific instructional context (Dorfner et al., 2017; Lipowsky et al., 2009; Steffensky & Neuhaus, 2018). Such connections could also be described in the cross-sectional analysis. The relationships found between PCK, PK, and diagnostic activities can be explained with regard to the items used. In the video-based assessment tool DiKoBi *Assess*, used to measure pre-service teachers' diagnostic activities, six subject-specific dimensions were covered: (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*, and (6) *conceptual instruction*. While the PCK paper-pencil test covered three of these dimensions (*student errors*, *use of models*, and *use of experiments*), there was also an overlap with two items of the PK test (*constructive handling of errors* and *role of students' prior knowledge*) (see Kramer, Förtsch, Boone et al., 2021). However, whereas the PK items asked for general strategies, the PCK items asked for content-related aspects such as specific student errors and their causes. Thus, they were closer to the video-based assessment tool. Such content-related concretization is necessary for subject-specific instruction (Dorfner et al., 2017; Steffensky et al., 2015). However, the results indicated that PK should not be disregarded in the context of subject-specific diagnostic competences, even if PCK has greater relevance.

4.4 Instructional support to foster knowledge development

For addressing the third aim of the dissertation, the two intervention studies were analyzed. Within biology education, these studies were one of the first ones that considered all three knowledge facets as conceptual part of separated or integrated instruction and as outcome variables of instruction. With regard to different ways of instructional support, both separated and integrated instructional support by a lecturer was yielding in terms of CK development. However, the descriptive results indicated potential benefits related to the separated instruction. As the correlations between the knowledge facets that were reported in science research suggest, CK is an important and necessary precondition for PCK (Fischer et al., 2012; Kleickmann et al., 2017). Teachers need a certain level of knowledge about subject-specific concepts and core ideas before they can effectively plan, teach, and reflect upon instruction (Ball et al., 2008; Kleickmann et al., 2017). This subject matter knowledge may best be taught in courses explicitly designed to build CK. A well-developed CK base can then be useful for integrating PCK and PK aspects, thus, developing integrated knowledge structures, which are more common among teachers with higher expertise (Berliner, 2001; Krauss et al., 2008). To verify this assumption explicitly, further research would have to be conducted in which effects of separated and integrated instruction are analyzed depending on the prior knowledge of the participants.

Regarding the PCK development, it was striking that not only the intervention treatments showed a significant increase in PCK, but also the control group, which only processed the knowledge tests and the video-based assessment tool DiKoBi *Assess*. This suggests that working on DiKoBi *Assess* also had a learning effect. Effects of the use of videos in terms of professional learning have already been observed in other studies (e.g., Gaudin & Chaliès, 2015; Llinares & Valls, 2009; Marsh et al., 2010; Roth et al., 2011). Videos that take into account the elements of cognitive load, student engagement, and active learning are particularly beneficial (Brame, 2016). Indicators of all three elements are applicable to the video-based assessment tool DiKoBi *Assess* (for example, important information was highlighted to reduce cognitive load, each video were kept brief (Min =.67 min; Max = 5.23 min) to keep student engagement high, and guiding questions have been implemented that might have been beneficial in terms of active learning). In addition, the situations shown in the videos represent prompts that can elicit knowledge (Alonzo & Kim, 2016; Kersting, 2008; Seidel & Stürmer, 2014), which in turn can influence further test performances. The effect resulting from the processing of the video-based tool might therefore have overshadowed possible interaction effects between time and treatments that should be induced by the intervention. However, treatments receiving separated or integrated instruction still had greater increases with larger effect sizes than the control group. That the greatest value of explained variance was found for the integrated instruction might indicate the potential benefit of addressing knowledge facets in an interrelated way in order to foster PCK development. This assumption is supported by similar findings from previous research

(Harr et al., 2014, 2015). A potential explanation for this is that PCK development can benefit from both the explicit instruction of PCK and the instruction of CK and PK aspects related to the specific content and situation (Kleickmann et al., 2017; Krauss et al., 2008; Tröbst et al., 2019). However, this assumption needs further investigation in the future.

Effects of instructional support by a lecturer on PK could not be analyzed in more detail due to ceiling effects that could be evaded in the future by increasing the level of the PK test (e.g., by adding higher cognitive test items or by using more test scales from the BilWiss project, see Kunina-Habenicht et al., 2020). However, a small time effect was found independent from treatments that may be due to the processing of DiKoBi *Assess* as well. Even though the videos showed situations in which the focus was on subject-specific instruction, the situated context also activated general pedagogical-psychological components that generally underlie instruction (e.g., knowledge of classroom management or knowledge of adaptivity and learning differences, see König et al., 2011; Wiens et al., 2020).

Instructional support via texts as it was implemented in Study 1 did not show measurable significant relevance for either professional knowledge or diagnostic activities. Although there was a significant increase in PCK and CK from pre to post-tests, this could not be attributed to the text-based intervention. As with instructional support by a lecturer, the video-based assessment tool might have caused such time effects. Consequently, the question arises as to how such influences of DiKoBi *Assess* can be explained. A possible explanation can be derived with reference to the RCM (see Section 1.3.1.1), which differentiates PCK into the realms cPCK (*collective*), pPCK (*personal*), and ePCK (*enacted*) (Carlson et al., 2019). Within the intervention studies, PCK was measured with paper-pencil tests eliciting pre-service teachers' declarative (knowing *that*) and action-related knowledge (knowing *how* and knowing *when and why*) (Förtsch et al., 2018). The majority of the items was presented as open-ended items and required pre-service teachers to write answers in their own words. It can be assumed that for solving the item tasks, both the specific education of the participants and their previous practical experiences were relevant. Consequently, the paper-pencil tests have measured participants' pPCK. In contrast, the instructional support (both the one by the lecturer and the one via texts) presented cPCK because they were designed on the basis of subject-specific literature. Additionally, the application of knowledge and diagnostic activities in the processing of DiKoBi *Assess* activated yet another realm of PCK connected to specific teaching actions (Alonzo et al., 2019). Prompted by the instructions given in the tool, participants engaged in scientific reasoning about the teacher's instructional behavior and instructional quality. The skills required are referred to as enacted PCK or ePCK (Carlson et al., 2019). Initial findings of other researchers supported the assumption that pPCK affects ePCK in terms of better (explaining) skills (Kulgemeyer et al., 2020). In addition, it is assumed that through reflection processes such as reasoning on teaching and learning situations, ePCK can, in turn, influence a teachers' pPCK (Alonzo et al., 2019). However, this assumption is difficult to prove due to the tacit and elusive nature of ePCK. Nevertheless, the results of the studies described in this dissertation suggest that ePCK may have affected pPCK through scientific reasoning about specific instructional situations (Alonzo et al., 2019). That might explain why PCK also increased in the control conditions. Thus, including practice-oriented tasks within teacher education, which elicit scientific reasoning processes, may be particularly beneficial for PCK, but potentially for PK as well.

Consequently, the question arises whether university courses, which mainly present knowledge separated in specialized faculties via texts and in lectures (Tröbst et al., 2019), could not endeavor to present knowledge, where possible and appropriate, in a more practice-oriented way. This might be especially beneficial for PCK and PK, as the acquisition and application of these knowledge facets is much more action-oriented and therefore requires additional forms of instructional support (cf. Kaiser et al., 2014; König et al., 2018). Consequently, the stronger integration of videos into university courses could be a corresponding element with which instruction on PK, CK, and PCK aspects could be applied in an interrelated way.

4.5 Limitations

Due to the multiple conceptual and methodological approaches that were applied in the individual studies of this dissertation, the limitations of the studies and analyses are discussed in terms of methodological and conceptual limitations below.

First, five limitations of methodological nature are stated. (1) The studies were conducted with a relatively small sample size due to their embedding into the regular courses of the biology education at LMU Munich. Therefore, the number of participants was limited. Because of the resulting lower statistical power, possible effects might not have been (clearly) visible. (2) Regarding the instructional support, the absence of manipulation checks can be criticized, as it could not be completely ensured that the interventions in the treatments were effectively processed. Certain precautions were taken (e.g., by preventing secondary activities such as distraction by mobile devices, and in turn, by using stimulating questions to enhance cognitive activation of the pre-service teachers), but processing was not instrumentally monitored. (3) Furthermore, in both intervention studies, the same paper-pencil tests were used for pre-testing and post-testing. Thus, a testing effect might have played a role in terms of a valid measurement (Shadish et al., 2002). It is also possible that the pre-service teachers acquired knowledge elsewhere during the period in which the respective study took place (several dates of data collection over 2-4 weeks) so that increases in knowledge cannot be explained exclusively by study effects. (4) In addition, overall analyses of the PK test showed that most of the selected items were too simple and therefore did not sufficiently discriminate the sample. In future studies, the PK test should be extended by using additional scales from the BilWiss project (Kunina-Habenicht et al., 2020) to increase test reliability (Linacre, 2021). (5) The video-based assessment tool DiKoBi *Assess*, which was primarily used as a measurement tool, emerged as potentially conducive in terms of learning. Explicit statistical evidence of this hypothesis does not yet exist, so future investigation of the tool as a learning environment has to follow. Regarding the videos from the video-based tool, it should be noted that these videos represented scripted, condensed challenges. In contrast, in real-life teaching, a multitude of complex challenges can occur. Consequently, measuring diagnostic competences with DiKoBi *Assess* may differ from measurements collected in real-life contexts. Nevertheless, the use of short videos of classroom situations in which the complexity of instructional practice is reduced is an important component for training inexperienced pre-service teachers' professional competence (Codreanu et al., 2020; Grossman et al., 2009).

Furthermore, five conceptual limitations are important to mention. (1) One conceptual limitation refers to the missing consideration of motivational variables in the analyses of this dissertation. The analyses presented in this dissertation focused on cognitive variables, whereas affective-motivational variables (such as interest or motivation) were collected separately and are available for later analyses. This procedure has proven to be successful for empirical research designs (Klieme & Leutner, 2006). However, since affective-motivational variables may have an impact on performance in problem- or case-based learning environments (Belland et al., 2013; Klug et al., 2016), they must also be considered when analyzing data collected with DiKoBi *Assess*. (2) For measuring diagnostic accuracy, pre-service teachers' descriptions of the identified challenges in DiKoBi *Assess* were used. The measure was calculated by comparing the instructional challenges described by the pre-service teachers with literature-supported challenges relevant for biology instruction. Thus, the calculation was related to the description of identified challenges so that diagnostic accuracy is more strongly associated with the diagnostic activity *generating evidence*. By doing so, the situated character of diagnostic decisions was taken into account. Nevertheless, other accuracy measures such as the development and embedding of separate items to measure diagnostic accuracy independent from diagnostic activities may be worth investigating in future studies. (3) Based on the validation, only those diagnostic activities that were considered relevant in the context of video analyses were considered for measurements and analyses with DiKoBi *Assess*. In real-life settings, other activities such as generating hypotheses may also be relevant (cf. Herppich et al., 2018; Wildgans-Lang et al., 2020). Furthermore, the comparison of the constructs *diagnostic activities* and *professional vision* presented as part of the validation was not completely tautology-free due to the specific study design and the data collected with it. However, the implemented diagnostic tasks can be considered highly relevant for effective diagnosing of video situations and with regard to real situations in biology classes. (4) The conceptualization of PK was rather narrow, covering merely pedagogical-psychological aspects of the dimension *instruction* that included the basic dimensions of instructional quality and general teaching methods. More comprehensive conceptualizations of PK, on the other hand, include even more dimensions, for example, knowledge of classroom assessment (König et al., 2011; Kunina-Habenicht et al., 2020; Voss & Kunter, 2013). An extension of the PK dimension would be particularly useful in relation to further research on instructional support in order to reflect the range of pedagogical course content. However, in the context of the research conducted with DiKoBi *Assess* and the focus on instructional quality, primarily the selected PK dimension was relevant. Finally, (5) professional knowledge was conceptualized in the present studies according to the common tripartite division into PK, CK, and PCK (cf. Baumert & Kunter, 2013a; Shulman, 1986, 1987). In the future, the knowledge facets could also be conceptualized more precisely, for example, with regard to the different realms of PCK defined in the RCM (cf. Carlson et al., 2019). That would be useful to investigate in more detail which PCK realms are promoted by which type of instructional support and to what extent there might be a difference between pre-service teachers with different prior knowledge.

4.6 Further research

Following up on the results of this dissertation, there are several aspects that require further research. Firstly, the use of an extended PK test version would be advisable to overcome the limitations of the present studies and to check whether results can be replicated. This would also be important in order to describe concepts for a stronger interlocking of PK and PCK courses and curricula (cf. König et al., 2018), and to be able to investigate their effectiveness by considering the various dimensions of PK and PCK included therein.

Further research could also pay more attention to pre-service teachers' prior knowledge by examining how different types of instructional support (e.g., different types of scaffolding) affect learning outcomes. This is currently being pursued in the follow-up project COSIMA 2, in which the video-based tool DiKoBi is used both as a measurement instrument (DiKoBi *Assess*) and as a learning environment (DiKoBi *Learn*). It is considered that "learners with low levels of prior knowledge might require more instructional support and guidance than advanced learners […], which therefore might influence the choice of instructional approaches" (Chernikova et al., 2020, p. 160). Building on the results of this dissertation regarding the importance of PCK for diagnostic competences in the context of subject-specific diagnosing, different scaffolding types for facilitating pre-service teachers' PCK will be investigated in COSIMA 2 for their effectiveness for pre-service biology teachers with different levels of prior knowledge. A recently published meta-analysis indicated that for learners with low prior knowledge providing examples is effective, whereas for learners with higher levels of prior knowledge, other scaffolding types such as prompts in the ongoing diagnostic process or the stimulation of reflection are more effective (Chernikova et al., 2020). The extent to which these findings can be translated into the promotion of PCK is currently being investigated.

While the diagnostic activities *generating evidence*, *evaluating evidence*, and *drawing conclusions* have been used as a joint measure for the analyses so far, it should also be investigated in the future to what extent different types of instructional support have different effects on the individual diagnostic activities and to what extent pre-service teachers at different stages of their education may need different forms of instructional support. Consequently, corresponding long-term studies are to be conducted here. In addition, the analyses should also be extended with regard to affective-motivational variables such as motivation or self-related cognitions that have a potential impact on teachers' skills as well (Kaiser et al., 2014; Klug et al., 2016; Meschede et al., 2017).

Furthermore, future studies within COSIMA should conceptually address differences in the measurement and promotion of different PCK realms. That requires a detailed decomposition and analysis of the measurement instruments and any intervention variables to intentionally conceptualize, control, and vary the different realms of PCK. Such efforts to understand the concept of PCK are currently being discussed mainly in the field of science teacher education and research (Sorge, Stender et al., 2019). The tools and instructional support approaches established in COSIMA can therefore be used to further evaluate the RCM to learn more about the exchanges between the broader knowledge facets and the three realms of PCK. Particularly "ePCK, i.e. how PCK is utilised in teachers' actual practice", still needs to be researched more intensively (Alonzo et al., 2019, p. 272). Investigating teachers' reasoning about instructional quality and their application of diagnostic activities might be a way to approximate this realm that is difficult to capture due to its tacit and elusive nature.

In addition, further research questions can be connected beyond the COSIMA project. Two contexts are pointed out in particular. First, the attempt to list different conceptualizations of situation-specific skills in order to compare study results revealed a high number of varying constructs (see Table 3 in Section 1.4). Even more, different labels exist for very similar conceptualizations. In order to bring an overview into the jungle of different constructs and to make it possible to compare or differentiate research results more precisely, a systematic review should bring structure into the research field. A similar concern was recently pursued by Santagata et al. (2021). However, their review referred to studies on mathematics teacher noticing that used video to support teacher learning and mainly addressed theoretical perspectives, the specific use of video technologies, and research questions and methods. A comparable approach was realized by Chan et al. (2020) for science education. But again, the focus of the review is on teacher noticing and professional vision. The future review should accordingly summarize and contrast conceptualizations of situation-specific skills in the context of science-specific classroom diagnosis and video analyses that go beyond teacher noticing and professional vision. Though, the comparison of *diagnostic activities* and *professional vision* presented in this dissertation can be used as a first step towards such an endeavor.

The second context refers to the extension of research designs toward classroom practice. Since diagnostic competences are relevant not only for video viewing but also in real-life instruction, future studies should advance more into investigating the instructional practice. How are diagnostic competences related to a teacher's instructional behavior and to instructional quality? It is also important to examine the extent to which diagnostic accuracy (in diagnosing subject-specific features of instructional quality) is related to performance measures of instructional quality. To date, the correlates of diagnostic accuracy with other measures are still not well enough studied (Kaiser et al., 2012; Machts et al., 2016; Schrader, 2013). The results of this dissertation explicitly indicate the relationship between PCK and (situated measures of) diagnostic accuracy, not only for student errors but also for other dimensions of subject-specific instructional quality. Whether teachers' diagnostic accuracy also turns out to be important for instructional quality in real-life teaching needs to be examined.

4.7 Implications

From the investigation of pre-service biology teachers' professional knowledge and diagnostic competences, some implications for research and practice can be derived. First, the present investigations refer to specific components of the COSIMA framework (Heitzmann et al., 2019; Chernikova et al., in press). Findings relate to the validation of the video-based assessment tool DiKoBi *Assess* to measure components of diagnostic competences (diagnostic activities and diagnostic accuracy), to relationships between components of diagnostic competences, and to the effects of instructional support provided by a lecturer or via texts. All findings were investigated within diagnostic situations relevant to biology instruction. Therefore, this dissertation provided answers to specific sections of this framework.

Regarding the interrelationships of components of diagnostic competences reported in this dissertation, practical implications for biology education can be derived. The promotion of pre-service teachers' PCK should primarily be addressed, for example, by teaching specific subject-related methods and principles and discussing them with regard to the subject-specific context (e.g., when using models or experiments) within didactical courses. The additional influence of PK found can be taken into account in these courses, for example, by drawing conclusions about general, broader principles when comparing and reflecting on specific scenarios (cf. Tröbst et al., 2019). This can be realized, for example, by examining a teacher's use of challenging tasks within different phases during the course of a biology lesson (e.g., in the reactivation, elaboration, and transfer phase) (cf. Dorfner, Förtsch, Spangler et al., 2019). In each phase, the method of stimulation of students to engage in higher-level thinking may differ (e.g., through guided instructional discussion, through specific work assignments), yet the concept of cognitive activation, which is one of the three basic dimensions of domain-general characteristics of instructional quality, overarches all observations (Lipowsky et al., 2009). Especially with regard to the instruction of knowledge about instructional quality, a stronger interlocking of PK and PCK thus makes sense.

Moreover, implications can be derived regarding the structure and methodology of university education. Through which instructional pathways can the development of knowledge and skills best be fostered? How can different faculties in teacher education cooperate and contribute to curricular cross-linking (cf. Meier et al., 2018)? Results from the separated or integrated instruction of knowledge facets provide further information regarding this question. As already emphasized by various researchers, a sound CK basis is considered crucial for the development of PCK (e.g., Baumert & Kunter, 2013b; Fischer et al., 2012). PCK, in turn, is considered crucial for the application of skills such as diagnostic activities in subject-specific contexts (cf. Hoth et al., 2016). Specialized courses that provide specific instruction on CK are therefore important. The present results of the dissertation support the importance of courses in which CK is taught separately. Nevertheless, CK can also benefit from integrated instruction of the knowledge facets. However, since such a rearrangement of university curricula is difficult to do (cf. Harr et al., 2015), separated instruction may be considered feasible and effective for CK development. For PCK, both separated and integrated instruction has been beneficial, with some evidence pointing to a small advantage for integrated instruction. Integration can be realized through linking PK, CK, and PCK aspects during course work. An example can be given with regard to the handling of student (mis)conceptions. First, general pedagogical-psychological principles of teaching and learning can be addressed (PK), which are then extended to student (mis)conceptions of a specific subject content (CK) by discussing how to deal with them subject-specifically (PCK). Overall, the research results suggest a stronger interlocking of didactical and pedagogical courses. Both the interrelated instruction of PCK and PK, as well as their application to specific contexts, can be expanded in future teacher education and should also be stated accordingly in the curricula. Instruction that connects PK and PCK aspects could be implemented in terms of more practical approaches. The results of this dissertation suggest that the sole focus on text- or lecture-based instruction is not the most optimal way, but that the development of the action-oriented knowledge facets PK and PCK can also be supported through other ways that allow a stronger interconnection of knowledge facets (Meier et al., 2018). Using videos or video-based tools can be considered promising in this regard. Using these tools in teacher education provides effective benefits (e.g., Kersting, 2008; Meschede et al., 2017; Roth et al., 2011; Seidel & Stürmer, 2014; Wiens et al., 2020): (1) videos elicit aspects of professional knowledge, (2) the situated context supports the acquisition and the application of knowledge and skills, (3) PK and PCK can be applied interrelatedly to the video cases, and (4) diagnostic competences can be acquired and trained with regard to specific, subject-relevant cases.

DiKoBi *Assess* (and in the future DiKoBi *Learn*) is a video-based tool that has been developed, validated, and used in initial studies to measure skills (i.e., diagnostic activities) within a situated diagnostic context. The tool provides subject-specific challenges from the biology classroom that are relevant for subject-specific instructional quality. In order to teach effectively and to implement specific features of instructional quality in the classroom, it is necessary that teachers are able to notice and diagnose relevant features (Meschede et al., 2017; Wiens et al., 2020). DiKoBi *Assess* can be used to measure this ability of teachers.

Using the tool systematically in the future can support pre-service biology teachers in their skills to notice relevant dimensions of instructional quality, analyze corresponding indicators, and develop alternative teaching strategies to improve biology-specific instruction. The use of DiKoBi *Assess* (and DiKoBi *Learn*) thus represents an important step in the promotion of pre-service biology teachers' knowledge application and its transfer to effective classroom instruction (cf. Jeschke et al., 2020; Seidel et al., 2017).

5. References

Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the Complexity of Science Teachers' PCK in Action: Enacted and Personal PCK. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (pp. 271–286). Springer.

Alonzo, A. C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, *53*(8), 1259–1286. https://doi.org/10.1002/tea.21271

Anderson, J. R. (1982). Acquisition of Cognitive Skill. *Psychological Review*, *89*(4), 369–406.

Arffman, I. (2016). Threats to validity when using open-ended items in international achievement studies: Coding responses to the PISA 2012 problem-solving test in Finland. *Scandinavian Journal of Educational Research*, *60*(6), 609–625. https://doi.org/10.1080/00313831.2015.1066429

Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften [Diagnostic Competence of Teachers]. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 157–160. https://doi.org/10.1024/1010-0652.23.34.157

Artelt, C., & Rausch, T. (2014). Accuracy of teacher judgments: When and for What Reasons? In S. Krolak-Schwerdt, S. Glock, & M. Böhmer (Eds.), *Teacher's professional development: Assessment, training, and learning. The future of education research* (pp. 27–43). Sense Publishers.

Ball, D. L. (2000). Bridging Practices: Intertwining Content and Pedagogy in Teaching and Learning to Teach. *Journal of Teacher Education,*, *51*(3), 241–247. https://doi.org/10.1177/0022487100051003013

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Bauer, E., Fischer, F., Kiesewetter, J., Shaffer, D. W., Fischer, M. R., Zottmann, J. M., & Sailer, M. (2020). Diagnostic Activities and Diagnostic Practices in Medical Education and Teacher Education. *Frontiers in Psychology*. Advance online publication. https://doi.org/10.3389/fpsyg.2020.562665

Baumert, J., & Kunter, M. (2013a). The COACTIV Model of Teachers' Professional Competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–48). Springer.

Baumert, J., & Kunter, M. (2013b). The Effect of Content Knowledge and Pedagogical Content Knowledge on Instructional Quality and Student Achievement. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 175–205). Springer.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Behling, F., Förtsch, C., & Neuhaus, B. J. (2019). 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]. *Zeitschrift für Didaktik der Naturwissenschaften*, *42*(2), 271. https://doi.org/10.1007/s40573-019-00103-9

Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A Framework for Designing Scaffolds That Improve Motivation and Cognition. *Educational Psychologist*, *48*(4), 243–270. https://doi.org/10.1080/00461520.2013.838920

Benedict-Chambers, A., & Aram, R. (2017). Tools for Teacher Noticing: Helping Preservice Teachers Notice and Analyze Student Thinking and Scientific Practice Use. *Journal of Science Teacher Education*, *28*(3), 294–318. https://doi.org/10.1080/1046560X.2017.1302730

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*, 463–482.

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, *121*, S2-S23. https://doi.org/10.1016/j.amjmed.2008.01.001

Blomberg, Geraldine, Renkl, A., Alexander, Sherin, G., Miriam, Borko, Hilda, Seidel, & Tina (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for Educational Research Online*, *5*(1), 90–114.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, *56*, 35–46. https://doi.org/10.1016/j.tate.2016.02.003

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies. *Zeitschrift für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Blömeke, S., & Kaiser, G. (2017). Understanding the Development of Teachers' Professional Competencies as Personally, Situationally and Socially Determined. In D. J. Clandinin & J. Husu (Eds.), *International handbook of research on teacher education* (pp. 783–802). Sage.

Blömeke, S., Kaiser, G., & Lehmann, R. (Eds.). (2010). *TEDS-M 2008: Professionelle Kompetenz und Lerngelegenheiten angehender Mathematiklehrkräfte für die Sekundarstufe I im internationalen Vergleich [TEDS-M 2008: Professional competence and learning opportunities of prospective secondary mathematics teachers in international comparison].* Waxmann.

Bond, T. G., & Fox, C. M. F. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum.

Boone, W. J., & Staver, J. (2020). *Advances in Rasch Analysis in the Human Science*. Springer.

Boone, W. J., Staver, J., & Yale, M. (2014). *Rasch analysis in the human sciences*. Springer.

Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-03-0125

Bromme, R. (1995). Was ist "pedagogical content knowledge"? Kritische Anmerkungen zu einem fruchtbaren Forschungsprogramm [What is „pedagogical content knowledge"? Critical comments on a fruitful research program]. *Zeitschrift für Pädagogik, Beiheft*, *33*, 105–113.

Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers [Competencies, functions and instructional behavior of the teacher]. In F. E. Weinert (Ed.), *Psychologie des Unterrichts und der Schule [Psychology of teaching and school]* (pp. 177–212). Hogrefe.

Brophy, J. (2000). *Teaching*. Educational Practices Series. http://www.ibe.unesco.org/fileadmin/user_upload/archive/Publications/educationalpracticesseriespdf/prac01e.pdf (accessed on 4 March 2021)

Brown, B. A., & Ryoo, K. (2008). Teaching science as a language: A "content-first" approach to science teaching. *Journal of Research in Science Teaching*, *45*(5), 529–553. https://doi.org/10.1002/tea.20255

Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning. *Educational Assessment*, *15*(3-4), 123–141. https://doi.org/10.1080/10627197.2010.530551

Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2013). The Diagnostic Skills of Mathematics Teachers. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 229–248). Springer.

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., . . . Wilson, C. D. (2019). The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (pp. 77–94). Springer.

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. C. (1988). Expert-Novice Differences in Perceiving and Processing Visual Classroom Information. *Journal of Teacher Education*, *39*, 25–31.

Cauet, E., Liepertz, S., Borowski, A., & Fischer, H. E. (2015). Does it Matter What We Measure? Domain-specific Professional Knowledge of Physics Teachers. *Swiss Journal of Educational Research*, *37*(3), 463–480. https://doi.org/10.24452/sjer.37.3.4963

Chan, K. K. H., Xu, L., Cooper, R., Berry, A., & Van Driel, J. (2020). Teacher noticing in science education: Do you see what I see? *Studies in Science Education*, *57*(2). https://doi.org/10.1080/03057267.2020.1755803

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2020). Facilitating Diagnostic Competences in Higher Education—a Meta-Analysis in Medical and Teacher Education. *Educational Psychology Review*, *32*, 157–196. https://doi.org/10.1007/s10648-019-09492-2

Chernikova, O., Heitzmann, N., Opitz, A., Seidel, T., & Fischer, F. (in press). A Theoretical Framework for Fostering Diagnostic Competences with Simulations. In F. Fischer & A. Opitz (Eds.), *Learning to Diagnose with Simulations - Examples from Teacher Education and Medical Education* (Chapter 2). Springer.

Codreanu, E., Sommerhoff, D., Huber, S., Ufer, S., & Seidel, T. (2020). Between authenticity and cognitive demand: Finding a balance in designing a video-based simulation in the context of mathematics teacher education. *Teaching and Teacher Education*, *95*. https://doi.org/10.1016/j.tate.2020.103146

Darling-Hammond, L. (2010). Teacher Education and the American Future. *Journal of Teacher Education*, *61*(1-2), 35–47. https://doi.org/10.1177/0022487109348024

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12–25. https://doi.org/10.1016/j.tate.2013.03.001

Dorfner, T., Förtsch, C., Germ, M., & Neuhaus, B. J. (2018). Biology instruction using a generic framework of scientific reasoning and argumentation. *Teaching and Teacher Education*, *75*, 232–243. https://doi.org/10.1016/j.tate.2018.07.003

Dorfner, T., Förtsch, C., & Neuhaus, B. J. (2017). 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]. *Zeitschrift für Didaktik der Naturwissenschaften*, *23*(1), 261–285. https://doi.org/10.1007/s40573-017-0058-3

Dorfner, T., Förtsch, C., & Neuhaus, B. J. (2018). Effects of three basic dimensions of instructional quality on students' situational interest in sixth-grade biology instruction. *Learning and Instruction*, *56*, 42–53. https://doi.org/10.1016/j.learninstruc.2018.03.001

Dorfner, T., Förtsch, C., & Neuhaus, B. J. (2019). Use of Technical Terms in German Biology Lessons and its Effects on Students' Conceptual Learning. *Research in Science & Technological Education*, *38*(2), 227–251. https://doi.org/10.1080/02635143.2019.1609436

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

Dreher, A., & Kuntze, S. (2015). Teachers' professional knowledge and noticing: The case of multiple representations in the mathematics classroom. *Educational Studies in Mathematics*, *88*(1), 89–114. https://doi.org/10.1007/s10649-014-9577-8

Evens, M., Elen, J., Larmuseau, C., & Depaepe, F. (2018). Promoting the development of teacher professional knowledge: Integrating content and pedagogy in teacher education. *Teaching and Teacher Education*, *75*, 244–258. https://doi.org/10.1016/j.tate.2018.07.001

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B. J., Dorner, B., Pankofer, S., Fischer, M. R., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *5*, 28–45. https://doi.org/10.14786/flr.v2i3.96

Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional Knowledge of Science Teachers. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second International Handbook of Science Education* (Vol. 85, pp. 435–448). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9041-7_30

Fischer, H. E., Labudde, P., Neumann, K., & Viiri, J. (2014). *Quality of instruction in physics*. Waxmann.

Fischer, J., Jansen, T., Möller, J., & Harms, U. (2021). Measuring biology trainee teachers' professional knowledge about evolution—introducing the Student Inventory. *Evolution: Education and Outreach*, *14*(1), 1105. https://doi.org/10.1186/s12052-021-00144-0

Fischer, J., Walter, J., Möller, J., & Harms, U. (2018). ProSim: Prozedurales Professionswissen im Simulierten Klassenraum [ProSim: Procedural Professional Knowledge in the Simulated Classroom]. IPN, CAU Kiel. https://www.e-teaching.org/etresources/pdf/best-practice_2018_fischer-walter-moeller-harms_neuesprosim-prozedurales-professionswissen-im-simulierten-klassenraum.pdf (accessed on 22 March 2021)

Förtsch, C., Dorfner, T., Baumgartner, J., Werner, S., Kotzebue, L. von, & Neuhaus, B. J. (2020). Fostering Students' Conceptual Knowledge in Biology in the Context of German National Education Standards. *Research in Science Education*, *50*, 739–771. https://doi.org/10.1007/s11165-018-9709-8

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M. R., Girwidz, R., Obersteiner, A., Reiss, K., Stürmer, K., Siebeck, M., Schmidmaier, R., Seidel, T., Ufer, S., Wecker, C., & Neuhaus, B. J. (2018). Systematizing Professional Knowledge of Medical Doctors and Teachers: Development of an Interdisciplinary Framework in the Context of Diagnostic Competences. *Education Sciences*, *8*(4), 207. https://doi.org/10.3390/educsci8040207

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

Förtsch, C., Werner, S., Kotzebue, L. von, & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education*, *38*(17), 2642–2666. https://doi.org/10.1080/09500693.2016.1257170

Furtak, E. M., Bakeman, R., & Buell, J. Y. (2018). Developing knowledge-in-action with a learning progression: Sequential analysis of teachers' questions and responses to student ideas. *Teaching and Teacher Education*, *76*, 267–282. https://doi.org/10.1016/j.tate.2018.06.001

Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., León, V. de, Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: findings from a four-year intervention study. *Instructional Science*, *44*(3), 267–291. https://doi.org/10.1007/s11251-016-9371-3

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

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

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

Glogger-Frey, I., Deutscher, M., & Renkl, A. (2018). Student teachers' prior knowledge as prerequisite to learn how to assess pupils' learning strategies. *Teaching and Teacher Education*, *76*, 227–241. https://doi.org/10.1016/j.tate.2018.01.012

Gold, B., & Holodynski, M. (2017). Using digital video to measure the professional vision of elementary classroom management: Test validation and methodological challenges. *Computers & Education*, *107*, 13–30. https://doi.org/10.1016/j.compedu.2016.12.012

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606–633.

Greeno, J. G. (2006). Theoretical and practical advances through research on learning. In J. L. Green, G. Camilli, P. B. Elmore, A. Skukauskaite, & E. Grace (Eds.), *Handbook of complementary methods in education research* (pp. 795–822). Lawrence Erlbaum.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching Practice: A Cross-Professional Perspective. *Teachers College Record*, *111*, 2055–2100.

Großschedl, J., Harms, U., Glowinski, I., & Waldmann, M. (2014). Professionswissen angehender Biologielehrkräfte: Das KiL-Projekt [Professional knowledge of prospective biology teachers: the KiL project]. *MNU (Mathematisch-Naturwissenschaftlicher Unterricht)*, *67*(8), 457–462.

Großschedl, J., Mahler, D., Kleickmann, T., & Harms, U. (2014). Content-Related Knowledge of Biology Teachers from Secondary Schools: Structure and learning opportunities. *International Journal of Science Education*, *36*(14), 2335–2366. https://doi.org/10.1080/09500693.2014.923949

Großschedl, J., Welter, V., & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, *56*(4), 402–439. https://doi.org/10.1002/tea.21482

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). *Building a science of classrooms: Application of the CLASS framework in over 4,000 U.S. early childhood and elementary classrooms.* https://www.icpsr.umich.edu/files/PREK3RD/resources/pdf/BuildingAScienceOfClassroomsPiantaHamre.pdf (accessed on 4 March 2021)

Harr, N., Eichler, A., & Renkl, A. (2014). Integrating pedagogical content knowledge and pedagogical/psychological knowledge in mathematics. *Frontiers in Psychology*, *5*, 1–10. https://doi.org/10.3389/fpsyg.2014.00924

Harr, N., Eichler, A., & Renkl, A. (2015). Integrated learning: ways of fostering the applicability of teachers' pedagogical and psychological knowledge. *Frontiers in Psychology*, *6*, 1–16. https://doi.org/10.3389/fpsyg.2015.00738

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., Schmidmaier, R., Neuhaus, B. J., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., & Fischer, F. (2019). Facilitating Diagnostic Competences in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education. *Frontline Learning Research*, *7*, 1–24. https://doi.org/10.14786/flr.v7i4.384

Helmke, A., & Lenske, G. (2013). Unterrichtsdiagnostik als Voraussetzung für Unterrichtsentwicklung [Diagnosis of Classroom Instruction from Different Perspectives as a Prerequisite for Improving Teaching and Learning]. *Beiträge zur Lehrerbildung*, *31*(2), 214–233.

Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education*, *3*(2), 91–98.

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Herppich, S., Praetorius, A.-K., Hetmanek, A., Glogger-Frey, I., Ufer, S., Leutner, D., Behrmann, L., Böhmer, I., Böhmer, M., Förster, N., Kaiser, J., Karing, C., Karst, K., Klug, J., Ohle, A., & Südkamp, A. (2017). Ein Arbeitsmodell für die empirische Erforschung der diagnostischen Kompetenz von Lehrkräften [A working model for empirical research on teachers‘ diagnostic competence]. In D. H. Rost (Ed.), *Diagnostische Kompetenz von Lehrkräften [Diagnostic Competence of Teachers]* (pp. 75–93). Waxmann.

Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2013). Benefits for processes cause decrements in outcomes: Training improves tutors‘ interactivity at the expense of assessment accuracy. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (Vol. 35, pp. 2530–2535). Cognitive Science Society. http://mindmodeling.org/cogsci2013/papers/0458/paper0458.pdf (accessed on 4 March 2021)

Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing Teachers to Learn from Teaching. *Journal of Teacher Education*, *58*(1), 47–61. https://doi.org/10.1177/0022487106295726

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, *42*(2), 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing Measures of Teachers' Mathematics Knowledge for Teaching. *The Elementary School Journal*, *105*(1), 11–30. https://doi.org/10.1086/428763

Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM*, *48*(1-2), 41–53. https://doi.org/10.1007/s11858-016-0759-y

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

Janssen, N., & Lazonder, A. W. (2016). Supporting pre-service teachers in designing technology-infused lesson plans. *Journal of Computer Assisted Learning*, *32*(5), 456–467. https://doi.org/10.1111/jcal.12146

Jentsch, A., Schlesinger, L., Heinrichs, H., Kaiser, G., König, J., & Blömeke, S. (2020). Erfassung der fachspezifischen Qualität von Mathematikunterricht: Faktorenstruktur und Zusammenhänge zur professionellen Kompetenz von Mathematiklehrpersonen [Measuring the subject-specific quality in mathematics instruction: factor structure and relations to mathematics teachers' professional competence]. *Journal für Mathematik-Didaktik*, *47*(1), 133. https://doi.org/10.1007/s13138-020-00168-x

Jeschke, C., Lindmeier, A., & Heinze, A. (2020). Vom Wissen zum Handeln: Vermittelt die Kompetenz zur Unterrichtsreflexion zwischen mathematischem Professionswissen und der Kompetenz zum Handeln im Mathematikunterricht? Eine Mediationsanalyse [From Knowledge to Action: Does the Competence to Prepare and Reflect on InstructionMediate Between Mathematics Teacher Knowledge and the Competence to Act in the Classroom? A Mediation Analysis]. *Journal für Mathematik-Didaktik.* Advance online publication. https://doi.org/10.1007/s13138-020-00171-2

Jüttner, M., Boone, W. J., Park, S., & Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educational Assessment, Evaluation and Accountability*, *25*(1), 45–67. https://doi.org/10.1007/s11092-013-9157-y

Jüttner, M., & Neuhaus, B. J. (2012). Development of Items for a Pedagogical Content Knowledge Test Based on Empirical Analysis of Pupils' Errors. *International Journal of Science Education*, *34*(7), 1125–1143. https://doi.org/10.1080/09500693.2011.606511

Jüttner, M., & Neuhaus, B. J. (2013). Validation of a Paper-and-Pencil Test Instrument Measuring Biology Teachers' Pedagogical Content Knowledge by Using Think-Aloud

Interviews. *Journal of Education and Training Studies*, *1*(2), 113–125. https://doi.org/10.11114/jets.v1i2.126

Kaiser, G., Blömeke, S., Busse, A., Döhrmann, M., & König, J. (2014). Professional knowledge of (prospective) Mathematics teachers – Its structure and development. In P. Liljedahl, C. Nicol, S. Oesterle, & Allan (Eds.), *Proceedings of the joint meeting of PME 38 and PME-NA 36* (pp. 35–50). PME.

Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers - cognitive versus situated approaches. *Educational Studies in Mathematics*, *94*(2), 161–184. https://doi.org/10.1007/s10649-016-9724-5

Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the Complexities of Video-Based Assessments: Theoretical and Methodological Approaches to Overcoming Shortcomings of Research on Teachers' Competence. *International Journal of Science and Mathematics Education*, *13*(2), 369–387. https://doi.org/10.1007/s10763-015-9616-7

Kaiser, J., Helm, F., Retelsdorf, J., Südkamp, A., & Möller, J. (2012). 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]. *Zeitschrift für Pädagogische Psychologie*, *26*(4), 251–261. https://doi.org/10.1024/1010-0652/a000076

Karing, C., Pfost, M., & Artelt, C. (2011). Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? [Are secondary teachers' diagnostic skills related to the development of their students' reading literacy and mathematical literacy?] *Journal for Educational Research Online*, *3*(2), 119–147.

Karst, K., Schoreit, E., & Lipowsky, F. (2014). 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]. *Zeitschrift für Pädagogische Psychologie*, *28*(4), 237–248. https://doi.org/10.1024/1010-0652/a000133

Kersting, N. B. (2008). Using Video Clips of Mathematics Classroom Instruction as Item Prompts to Measure Teachers' Knowledge of Teaching Mathematics. *Educational and Psychological Measurement*, *68*(5), 845–861. https://doi.org/10.1177/0013164407313369

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring Usable Knowledge: Teachers' Analyses of Mathematics Classroom Videos Predict Teaching Quality and Student Learning. *American Educational Research Journal*, *49*(3), 568–589.

Kirschner, S., Borowski, A., Fischer, H. E., Gess-Newsome, J., & Aufschnaiter, C. von (2016). Developing and evaluating a paper-and-pencil test to assess components of physics teachers' pedagogical content knowledge. *International Journal of Science Education*, *38*(8), 1343–1372. https://doi.org/10.1080/09500693.2016.1190479

Kleickmann, T., Tröbst, S., Heinze, A., Bernholt, A., Rink, R., & Kunter, M. (2017). Teacher Knowledge Experiment: Conditions of the Development of Pedagogical Content Knowledge. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Methodology of Educational Measurement and Assessment. Competence assessment in education: Research, models and instruments* (Vol. 59, pp. 111–129). Springer. https://doi.org/10.1007/978-3-319-50030-0_8

Klieme, E., Hartig, J., & Rauch, D. (2008). Concept of Competence in Educational Contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 3–22). Hogrefe & Huber Publishers.

Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG [Competence models for assessing individual learning outcomes and balancing educational processes. Description of a newly established priority program of the DFG]. *Zeitschrift für Pädagogik*, *52*(6), 876–903.

Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts ‘‘Pythagoras’’ [Quality dimensions and effectiveness of mathematics instruction. Theoretical background and selected findings of the Pythagoras project]. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (pp. 127–146). Waxmann.

Klieme, E., & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts [Empirical classroom research and subject didactics. Outcome-oriented measurement and process quality of teaching]. *Zeitschrift für Pädagogik*, *54*, 2, 222-237.

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

Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education*, *30*, 38–46. https://doi.org/10.1016/j.tate.2012.10.004

Klug, J., Bruder, S., & Schmitz, B. (2016). Which variables predict teachers diagnostic competence when diagnosing students’ learning behavior at different stages of a teacher’s career? *Teachers and Teaching*, *22*(4), 461–484. https://doi.org/10.1080/13540602.2015.1082729

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers’ general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, *38*, 76–88. https://doi.org/10.1016/j.tate.2013.11.004

König, J., Blömeke, S., Paine, L., Schmidt, W. H., & Hsieh, F.-J. (2011). General Pedagogical Knowledge of Future Middle School Teachers: On the Complex Ecology of Teacher Education in the United States, Germany, and Taiwan. *Journal of Teacher Education*, *62*(2), 188–201. https://doi.org/10.1177/0022487110388664

König, J., Doll, J., Buchholtz, N., Förster, S., Kaspar, K., Rühl, A.-M., Strauß, S., Bremerich-Vos, A., Fladung, I., & Kaiser, G. (2018). Pädagogisches Wissen versus fachdidaktisches Wissen? [General pedagogical knowledge versus pedagogical content knowledge?]. *Zeitschrift für Erziehungswissenschaft*, *21*(3), 1–38. https://doi.org/10.1007/s11618-017-0765-z

König, J., & Kramer, C. (2016). Teacher professional knowledge and classroom management: on the relation of general pedagogical knowledge (GPK) and classroom management expertise (CME). *ZDM*, *48*(1-2), 139–151. https://doi.org/10.1007/s11858-015-0705-4

Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computersupported, case-based worked examples: effects of erroneous examples and feedback. *Medical Education*, *42*, 823–829. https://doi.org/10.1111/j.1365-2923.2008.03122.x

Kramer, M., Förtsch, C., Boone, W. J., Seidel, T., & Neuhaus, B. J. (2021). Investigating Pre-Service Biology Teachers' Diagnostic Competence: Systematic Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy. *Education Sciences*, *11*(3), 89. https://doi.org/10.3390/educsci11030089

Kramer, M., Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation*, *28.* https://doi.org/10.1016/j.stueduc.2020.100973

Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, *7*(1). https://doi.org/10.1080/2331186X.2020.1823155

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716–725. https://doi.org/10.1037/0022-0663.100.3.716

Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M., Jordan, A., & Löwen, K. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz [Teachers' professional knowledge, cognitively activating mathematics instruction, and the development of mathematical competence]. In J. Doll & M. Prenzel (Eds.), *Bildungsqualitätvon Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung [Educational quality in schools: teacher professionalization, instructional development, and student support as strategies for quality improvement]* (pp. 31–53). Waxmann.

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

Kulgemeyer, C., Borowski, A., Buschhüter, D., Enkrott, P., Kempin, M., Reinhold, P., Riese, J., Schecker, H., Schröder, J., & Vogelsang, C. (2020). Professional knowledge affects action-related skills: The development of preservice physics teachers' explaining skills during a field experience. *Journal of Research in Science Teaching*, *4*(3), 1105. https://doi.org/10.1002/tea.21632

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

Kunter, M., & Baumert, J. (2013). The COACTIV Research Program on Teachers' Professional Competence: Summary and Discussion. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 345–368). Springer.

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

Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., Krauss, S., Löwen, K., Neubrand, M., & Tsai, Y.-M. (2007). Linking Aspects of Teacher Competence to their Instruction: Results from the COACTIV Project. In M. Prenzel (Ed.), *Studies on the educational quality of schools: The final report on the DFG Priority Programme* (pp. 32–52). Waxmann.

Kunter, M., Leutner, D., Terhart, E., & Baumert, J. (2014). *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]*. IQB – Institut zur Qualitätsentwicklung im Bildungswesen. https://www.iqb.hu-berlin.de/fdz/studies/BilWiss

Lai, M. K., & Schildkamp, K. (2016). In-service Teacher Professional Learning: Use of Assessment in Data-based Decision-making. In G. Brown (Ed.), *Handbook of Human and Social Conditions in Assessment* (pp. 77–94). Routledge.

Lazonder, A. W., & Harmsten, R. (2016). Meta-Analysis of Inquiry-Based Learning: Effects of Guidance. *Review of Educational Research*, *86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Lee, E., & Luft, J. A. (2008). Experienced Secondary Science Teachers' Representation of Pedagogical Content Knowledge. *International Journal of Science Education*, *30*(10), 1343–1363. https://doi.org/10.1080/09500690802187058

Lehane, L., & Bertram, A. (2016). Getting to the CoRe of it: A review of a specific PCK conceptual lens in science educational research. *Educación Química*, *27*(1), 52–58. https://doi.org/10.1016/j.eq.2015.09.004

Lenske, G., Wagner, W., Wirth, J., Thillmann, H., Cauet, E., Liepertz, S., & Leutner, D. (2016). Die Bedeutung des pädagogisch-psychologischen Wissens für die Qualität der Klassenführung und den Lernzuwachs der Schüler/innen im Physikunterricht [The importance of pedagogical knowledge for classroom management and for students' achievement]. *Zeitschrift für Erziehungswissenschaft*, *19*(1), 211–233. https://doi.org/10.1007/s11618-015-0659-x

Linacre, J. M. (2021). *A User's Guide to Winsteps® Ministeps Rasch-Model Computer Programs: Program Manual 4.8.0*. https://www.winsteps.com/a/Winsteps-Manual.pdf (accessed on 4 March 2021)

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, *19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Llinares, S., & Valls, J. (2009). The building of pre-service primary teachers' knowledge of mathematics teaching: interaction and online video case studies. *Instructional Science*, *37*(3), 247–271. https://doi.org/10.1007/s11251-007-9043-4

Loibl, K., Leuders, T., & Dörfler, T. (2020). A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teaching and Teacher Education*, *91*(3), 103059. https://doi.org/10.1016/j.tate.2020.103059

Machts, N., Kaiser, J., Schmidt, F. T.C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, *19*, 85–103. https://doi.org/10.1016/j.edurev.2016.06.003

Marsh, B., Mitchell, N., & Adamczyk, P. (2010). Interactive video technology: Enhancing professional learning in initial teacher education. *Computers & Education*, *54*(3), 742–748. https://doi.org/10.1016/j.compedu.2009.09.011

Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen [Knowledge acquisition as scientific problem solving]. In D. Krüger & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden [Theories in biology education research. A handbook for pre-service teachers and doctoral students]* (pp. 177–186). Springer.

McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften [Diagnostic skills of teachers]. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 223–235. https://doi.org/10.1024/1010-0652.23.34.223

Meier, M., Grospietsch, F., & Mayer, J. (2018). Vernetzung von Wissensfacetten professioneller Handlungskompetenz in hochshuldidaktischen Lehr-Lernsettings [Linking of knowledge facets of professional acting in teaching-learning settings of higher education]. In I. Glowinski, A. Borowski, J. Gillen, S. Schanze, & J. von Meien (Eds.), *Kohärenz in der universitären Lehrerbildung [Coherence in university teacher education]* (pp. 143–178). Universitätsverlag Potsdam.

Meschede, N., Fiebranz, A., Möller, K., & Steffensky, M. (2017). Teachers' professional vision, pedagogical content knowledge and beliefs: On its relation and differences between pre-service and in-service teachers. *Teaching and Teacher Education*, *66*, 158–170. https://doi.org/10.1016/j.tate.2017.04.010

Michalsky, T. (2014). Developing the SRL-PV assessment scheme: Preservice teachers' professional vision for teaching self-regulated learning. *Studies in Educational Evaluation*, *43*, 214–229. https://doi.org/10.1016/j.stueduc.2014.05.003

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press. https://doi.org/10.17226/13165

Nawani, J., Kotzebue, L., Rixius, J., Graml, M., & Neuhaus, B. J. (2017). Teachers' Use of Focus Questions in German Biology Classrooms: a Video-based Naturalistic Study. *International Journal of Science and Mathematics Education*, *95*(4), 639. https://doi.org/10.1007/s10763-017-9837-z

Neuhaus, B. J. (2007). Unterrichtsqualität als Forschungsfeld für empirische biologiedidaktische Studien [Instructional quality as research field for empirical studies in biology education]. In D. Krüger & H. Vogt (Eds.), *Theorien in der biologiedidaktischen Forschung: Ein Handbuch für Lehramtsstudenten und Doktoranden [Theories in biology education research. A handbook for pre-service teachers and doctoral students]* (pp. 143–154). Springer.

Oh, P. S., & Oh, S. J. (2011). What Teachers of Science Need to Know about Models: An overview. *International Journal of Science Education*, *33*(8), 1109–1130. https://doi.org/10.1080/09500693.2010.502191

Ohle, A., & McElvany, N. (2015). Teacher diagnostic competences and their practical relevance. Special issue editorial. *Journal for Educational Research Online*, *7*(2), 5–10.

Ostermann, A., Leuders, T., & Nückles, M. (2018). Improving the judgment of task difficulties: prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education*, *21*(6), 579–605. https://doi.org/10.1007/s10857-017-9369-z

Ostermann, A., Leuders, T., & Philipp, K. (2019). Fachbezogene diagnostische Kompetenzen von Lehrkräften – Von Verfahren der Erfassung zu kognitiven Modellen zur Erklärung [Subject-related diagnostic competencies of teachers - From methods of assessment to cognitive models of explanation]. In T. Leuders, M. Nückles, S. Mikelskis-Seifert, & K. Philipp (Eds.), *Pädagogische Professionalität in Mathematik und Naturwissenschaften*

[Pedagogical professionalism in mathematics and science] (pp. 93–116). Springer Spektrum.

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a Strategic Reader. *Contemporary Educational Psychology*, *8*, 293–316.

Park, S., & Oliver, J. S. (2008). Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Research in Science Education*, *38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Pianta, R. C., Hamre, B. K., & Mintz, S. L. (2012). *Classroom assessment scoring system (CLASS): Secondary class manual*. Teachstone.

Pickal, A., Wecker, C., Neuhaus, B. J., & Girwidz, R. (in press). Learning to Diagnose Secondary School Students' Scientific Reasoning Skills in Physics and Biology: Video-Based Simulations for Pre-Service Teachers. In F. Fischer & A. Opitz (Eds.), *Learning to Diagnose with Simulations - Examples from Teacher Education and Medical Education* (chapter 7). Springer.

Piwowar, V., Barth, V. L., Ophardt, D., & Thiel, F. (2018). Evidence-based scripted videos on handling student misbehavior: the development and evaluation of video cases for teacher education. *Professional Development in Education*, *44*(3), 369–384. https://doi.org/10.1080/19415257.2017.1316299

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*, *50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Praetorius, A.-K., & Südkamp, A. (2017). Eine Einführung in das Thema der diagnostischen Kompetenz von Lehrkräften [An introduction to the topic of teachers' diagnostic competence]. In D. H. Rost (Ed.), *Diagnostische Kompetenz von Lehrkräften [Diagnostic Competence of Teachers]* (pp. 13–18). Waxmann.

EFS Survey (Version Winter 2018) [Computer software]. (2018). Questback GmbH. Köln.

Rach, S., Ufer, S., & Heinze, A. (2013). Learning from errors: Effects of teachers' training on students' attitudes towards and their individual use of errors. *Proceedings of the National Academy of Sciences*, *8*(1), 21–30.

Rakoczy, K., Klieme, E., Leiß, D., & Blum, W. (2017). Formative Assessment in Mathematics Instruction: Theoretical Considerations and Empirical Results of the Co2CA Project. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Methodology of Educational Measurement and Assessment. Competence assessment in education: Research, models and instruments* (pp. 447–467). Springer.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. *Educational Psychologist*, *38*(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert Knowledge: Analyses and Remedies. *Educational Psychologist*, *31*(2), 115–121.

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I.Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, *48*(2), 117–148. https://doi.org/10.1002/tea.20408

Santagata, R., König, J., Scheiner, T., Nguyen, H., Adleff, A.-K., Yang, X., & Kaiser, G. (2021). Mathematics teacher learning to notice: a systematic review of studies of video-based programs. *ZDM – Mathematics Education*, *13*(3), 223. https://doi.org/10.1007/s11858-020-01216-z

Santagata, R., & Yeh, C. (2016). The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM*, *48*(1-2), 153–165. https://doi.org/10.1007/s11858-015-0737-9

Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: an empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, *10*(2), 123–140. https://doi.org/10.1007/s10857-007-9029-9

Sauvé, L., Renaud, L., Kaufman, D., & Marquis, J.-S. (2007). Distinguishing between games and simulations: A systematic review. *Educational Technology & Society*, *10*(3), 247–256.

Schäfer, S., & Seidel, T. (2015). Noticing and reasoning of teaching and learning components by pre-service teachers. *Journal for Educational Research Online*, *7*(2), 34–58.

Scheiner, T. (2016). Teacher noticing: enlightening or blinding? *ZDM*, *48*(1-2), 227–238. https://doi.org/10.1007/s11858-016-0771-2

Schiefele, U. (2009). Situational and Individual Interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of Motivation at School* (197-222). Routledge.

Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM*, *48*(1-2), 29–40. https://doi.org/10.1007/s11858-016-0765-0

Schmelzing, S., Van Driel, J., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the 'cardiovascular system'. *International Journal of Science and Mathematics Education*, *11*(6), 1369–1390.

Schmidt, W. H., Tatto Maria T., Bankov, K., Blömeke, S., Cedillo, T., Cogan, L., Han, S. I., Houang, R., Hsieh, F. J., Paine, L., Santillan, M., & Schwille, J. (2007). *The preparing gap: Teacher education for middle school mathematics in six countries (MT21 Report).* MSU Center for Research in Mathematics and Science Education. http://www.educ.msu.edu/content/sites/usteds/documents/MT21Report.pdf (accessed on 4 March 2021)

Schrader, F.-W. (1989). *Diagnostische Kompetenz von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts [Diagnostic competence of teachers and its importance for the design and effectiveness of teaching]*. Peter Lang.

Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen [Teacher Diagnosis and Diagnostic Competence]. *Beiträge Zur Lehrerbildung*, *31*(2).

Schwartz, A., & Elstein, A. S. (2008). Clinicial reasoning in medicine. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 223–234). Elsevier (Butterworth Heinemann).

Scott, P., Mortimer, E., & Ametller, J. (2011). Pedagogical link-making: a fundamental aspect of teaching and learning scientific conceptual knowledge. *Studies in Science Education*, *47*(1), 3–36. https://doi.org/10.1080/03057267.2011.549619

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer" - Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE ["Observer" - Validation of a video-based instrument to capture professional vision of teaching. Project OBSERVE]. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Zeitschrift für Pädagogik, Beiheft: Vol. 56. Kompetenzmodellierung: Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes[Competence Modeling: Interim Review of the DFG Priority Program and Perspectives of the Research Approach]* (pp. 296–306). Beltz.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, *77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, *51*(4), 739–771. https://doi.org/10.3102/0002831214531321

Seidel, T., Stürmer, K., Prenzel, M., Jahn, G., & Schäfer, S. (2017). Investigating Pre-service Teachers'v Professional Vision Within University-Based Teacher Education. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Methodology of Educational Measurement and Assessment. Competence assessment in education: Research, models and instruments* (pp. 93–109). Springer. https://doi.org/10.1007/978-3-319-50030-0_7

Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2008). Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung [Common state requirements for the subject sciences and subject didactics in teacher education]. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2008/2008_10_16-Fachprofile-Lehrerbildung.pdf (accessed on 4 March 2021)

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Sherin, M. G., & van Es, E. A. (2009). Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education*, *60*(1), 20–37. https://doi.org/10.1177/0022487108328155

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

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Havard Educational Review*, *57*, 1–22.

Simon, H. A. (1992). Scientific discovery as problem solving. *International Studies in Philosophy of Science*, *6*(1), 3-14.

Sorge, S., Keller, M. M., Neumann, K., & Möller, J. (2019). Investigating the relationship between pre-service physics teachers' professional knowledge, self-concept, and interest. *Journal of Research in Science Teaching*, *56*(7), 937–955. https://doi.org/10.1002/tea.21534

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

Spychiger, M. (2008). Lernen aus Fehlern und Entwicklung von Fehlerkultur. Konzeptuelle Grundlagen und programmatische Thesen für einen pädagogischen Umgang mit Fehlern [Learning from errors and developing an error culture. Conceptual foundations and programmatic theses for a pedagogical approach to errors]. *Erwägen Wissen Ethik (Streitforum für Erwägungskultur)*, *19*(3), 274-282.

Stahnke, R., Schueler, S., & Roesken-Winter, B. (2016). Teachers' perception, interpretation, and decision-making: a systematic review of empirical mathematics education research. *ZDM*, *48*(1-2), 1–27. https://doi.org/10.1007/s11858-016-0775-y

Star, J. R., & Strickland, S. K. (2008). Learning to observe: using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education*, *11*(2), 107–125. https://doi.org/10.1007/s10857-007-9063-7

State Institute of School Quality and Educational Research, Munich. (2017). LehrplanPLUS Mittelschule: Natur und Technik 5. https://www.lehrplanplus.bayern.de/fachlehrplan/mittelschule/5/nt/regelklasse#77436 (accessed on 20 March 2021)

Steffensky, M., Gold, B., Holdynski, M., & Möller, K. (2015). Professional Vision of Classroom Management and Learning Support in Science Classrooms—Does Professional Vision Differ Across General and Content-Specific Classroom Interactions? *International Journal of Science and Mathematics Education*, *13*(2), 351–368. https://doi.org/10.1007/s10763-014-9607-0

Steffensky, M., & Neuhaus, B. J. (2018). Unterrichtsqualität im naturwissenschaftlichen Unterricht [Instructional quality in science teaching]. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Theorien in der naturwissenschaftsdidaktischen Forschung [Theories in science education research]*. (Vol. 47, pp. 299–313). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56320-5_18

Stürmer, K., Könings, K. D., & Seidel, T. (2012). Declarative knowledge and professional vision in teacher education: Effect of courses in teaching and learning. *The British Journal of Educational Psychology*, *83*(3), 467–483. https://doi.org/10.1111/j.2044-8279.2012.02075.x

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*, 743–762.

Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der simulierte Klassenraum [The simulated classroom]. *Zeitschrift für Pädagogische Psychologie*, *22*(3-4), 261–276. https://doi.org/10.1024/1010-0652.22.34.261

Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., Jüttner, M., Kirschner, S., Leutner, D., Neuhaus, B. J., Sandmann, A., Sumfleth, E., Hubertina, T., & Wirth, J. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften [Item Development Model for Assessing Professional Knowledge of Science Teachers]. *Zeitschrift für Didaktik der Naturwissenschaften*, *18*.

Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie [Experimentation in physics lessons - results of a video study]. *Zeitschrift für Didaktik der Naturwissenschaften*, *10*, 51–69.

Tolsdorf, Y., & Markic, S. (2017). Exploring Chemistry Student Teachers' Diagnostic Competence—A Qualitative Cross-Level Study. *Education Sciences*, *7*(4), 86. https://doi.org/10.3390/educsci7040086

Tröbst, S., Kleickmann, T., Depaepe, F., Heinze, A., & Kunter, M. (2019). Effects of instruction on pedagogical content knowledge about fractions in sixth-grade mathematics on content knowledge and pedagogical knowledge. *Unterrichtswissenschaft*, *47*(1), 79–97. https://doi.org/10.1007/s42010-019-00041-y

Ufer, S., & Leutner, D. (2017). Kompetenzen als Dispositionen - Begriffserklärungen und Herausforderungen [Competences as Dispositions - Explanations of Terms and Challenges]. In D. H. Rost (Ed.), *Diagnostische Kompetenz von Lehrkräften [Diagnostic Competence of Teachers]* (pp. 67–74). Waxmann.

van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher*, *10*(4), 571–596.

van Lehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*(1), 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

van Ophuysen, S. (2010). Professionelle pädagogisch-diagnostische Kompetenz. Eine theoretische und empirische Annäherung [Professional pedagogical-diagnostic competence. A theoretical and empirical approach]. *Jahrbuch der Schulentwicklung*, *16*, 203–234.

van Ophuysen, S., & Behrmann, L. (2015). Die Qualität pädagogischer Diagnostik im Lehrerberuf - Anmerkungen zum Themenheft „Diagnostische Kompetenzen von Lehrkräften und ihre Handlungsrelevanz“ [The Quality of Pedagogical Diagnostics in the Teaching Profession - Comments on the Topic Booklet “Diagnostic Competences of Teachers and their Relevance for Practice“]. *Journal for Educational Research Online*, *7*(2), 82–98.

Vogt, F., & Schmiemann, P. (2020). Assessing Biology Pre-Service Teachers’ Professional Vision of Teaching Scientific Inquiry. *Education Sciences*, *10*(11), 332. https://doi.org/10.3390/educsci10110332

Voss, T., & Kunter, M. (2013). Teachers’ General Pedagogical/Psychological Knowledge. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 207–227). Springer. https://doi.org/10.1007/978-1-4614-5149-5_10

Wadouh, J., Liu, N., Sandmann, A., & Neuhaus, B. J. (2014). The Effect of Knowledge linking Levels in Biology Lessons upon Students’ Knowledge Structure. *International Journal of Science and Mathematics Education*, *12*, 25–47. https://doi.org/10.1007/s10763-012-9390-8

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

Weinert, F. E., Schrader, F.-W., & Helmke, A. (1990). Educational expertise. Closing the gap between educational research and classroom practice. *School Psychology International*, *11*, 163–180.

Werner, S., Förtsch, C., Boone, W. J., Kotzebue, L. von, & Neuhaus, B. J. (2017). Investigating how german biology teachers use three-dimensional physical models in classroom instruction: A video study. *Research in Science Education*, *1*(2), 195. https://doi.org/10.1007/s11165-017-9624-4

Wiedmann, M., Kaendler, C., Leuders, T., Spada, H., & Rummel, N. (2019). Measuring teachers’ competence to monitor student interaction in collaborative learning settings. Unterrichtswissenschaft, 47(2), 177–199. https://doi.org/10.1007/s42010-019-00047-6

Wiens, P. D., Beck, J. S., & Lunsmann, C. J. (2020). Assessing teacher pedagogical knowledge: the Video Assessment of Teacher Knowledge (VATK). *Educational Studies*, 1–17. https://doi.org/10.1080/03055698.2020.1750350

Wildgans-Lang, A., Scheuerer, S., Obersteiner, A., Fischer, F., & Reiss, K. (2020). Analyzing prospective mathematics teachers’ diagnostic processes in a simulated environment. *ZDM*, *57*(1), 175. https://doi.org/10.1007/s11858-020-01139-9

Wirtz, M. A., & Caspar, F. (2002). *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.

Wüsten, S. (2010). *Allgemeine und fachspezifische Merkmale der Unterrichtsqualität im Fach Biologie: Eine Video- und Interventionsstudie [General and subject-specific instructional quality in the subject biology. A video and intervention study]* [Dissertation]. Universität Duisburg-Essen.

6. Acknowledgments

I would like to express my sincere gratitude to Birgit J. Neuhaus for supervising my thesis and offering me the doctoral position in the COSIMA project. Taking part in the scientific research community gave me various opportunities to learn, grow, laugh, interact, and communicate, but also to experience time of trouble and stress without letting it get me down. I believe that this inspiring research community has contributed a great deal to my personal development. Without your promotional words, Mrs. Neuhaus, I would never have been able to enjoy this community. In addition, you were always on call when there were problems to be solved, even at very short notice, which greatly benefited the orderly progress of all kinds of my projects. In retrospect, I can say that you also found a good balance between letting me walk on a long leash to pursue my own ideas and never letting me so far out of sight that I might have taken a wrong turn.

A second big smile goes to Christian Förtsch. Right from the start, you were a tremendous support to me when studies had to be planned, theories had to be transferred, or articles had to be revised. Your ideas and problem-solving strategies were always right on point. You supported me in all matters and were always the first contact person for me. When you finally accepted that I simply do not like beer and nevertheless continued to give me your full support, it was a real weight off my mind.

Moreover, I would like to express my deepest thanks to the DFG research group COSIMA and all professors, post-docs, and colleagues associated with this project. Explicitly, I want to thank Tina Seidel. With your research expertise, you precisely set important conceptual points on my path. In addition, your sincere, calm and supportive manner had an incredibly grounding effect and always made conversations delightful and enriching. Furthermore, a genuine thank you goes to my COSIMA colleagues Elias Codreanu, Anika Radkowitsch, Max Fink, Amadeus Pickal, Elisabeth Bauer, Angelika Wildgans-Lang, Christians Schons, Stephanie Kron, Constanze Richters, Michael Nickl, Marie Irmer, and Caroline Corves, with whom I not only removed mountains of exasperation but also reached for real mountains. It is always nice to know that you can also count on each other and have fun together privately.

Moreover, this thesis could not have been realized without the help of Julia Stürmer, Sonja Förtsch, William J. Boone, Martha Ledl, Leoni Bergmann, Elisabeth Steffl, and the support of all those people working in the lab of Birgit J. Neuhaus. Even during the Corona situation, a coffee chat via Zoom was never out of the question, a supportive message via Messenger never far away. So thanks to the entire AG for simply being there when help was needed. And many thanks to Nicole Methfessel, who handled all the bureaucratic tasks.

A final thank you belongs to the people from my closest circle of supporters. First of all, I thank my parents, to whom I could always retreat and be sure of good care. I thank my sister Luise, with whom I can experience adventures in the wide world and share moments that only exist among siblings and that give me so much strength. And last but not least, Robin Maier, who suddenly appeared in front of me one day during my doctoral studies and has not disappeared since, who goes through the darkest holes with me but also climbs the highest peaks, who keeps my energy level in balance with his culinary skills, who is cheering up personified.

Thank you!

7. Curriculum vitae

(*Hinweis*: Der Lebenslauf ist nur in der gebundenen Fassung der Dissertation abgedruckt.)

Publikationen – peer-reviewed Journals

Submitted

Kramer, M., Förtsch, C., & Neuhaus, B. J. (under revision). Can pre-service biology teachers' professional knowledge and diagnostic activities be fostered by self-directed knowledge acquisition via texts? *Education Sciences*.

Kramer, M., Förtsch, C., & Neuhaus, B. J. (moderate revision). Integrating or not-integrating – that is the question. Effects of integrated instruction on the development of pre-service biology teachers' professional knowledge. *Frontiers in Education*.

2021

Kramer, M., Förtsch, C., Boone, W. J., Seidel, T., & Neuhaus, B. J. (2021). Investigating Pre-Service Biology Teachers' Diagnostic Competences: Relationships between Professional Knowledge, Diagnostic Activities, and Diagnostic Accuracy. *Education Sciences*, *11*(3), 89. https://doi.org/10.3390/educsci11030089

Kramer, M., Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation*, *28*. https://doi.org/10.1016/j.stueduc.2020.100973

2020

Kramer, M., Förtsch, C., Stürmer, J., Förtsch, S., Seidel, T., & Neuhaus, B. J. (2020). Measuring biology teachers' professional vision: Development and validation of a video-based assessment tool. *Cogent Education*, *7*(1). https://doi.org/10.1080/2331186X.2020.1823155

2019

Kramer, M., Förtsch, C., Aufleger, M., & Neuhaus, B. J. (2019). Der Einsatz digitaler Medien im gymnasialen Biologieunterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, *25*(1), 131-160. https://doi.org/10.1007/s40573-019-00096-5

Beiträge in Herausgeberwerken

In press

Kramer, M., Stürmer, J., Förtsch, C., Seidel, T., Ufer, S., Fischer, M. R., & Neuhaus, B. J. (in press). Diagnosing the Instructional Quality of Biology Lessons Based on Staged Videos: Developing DiKoBi, a Video-Based Simulation. In F. Fischer & A. Opitz (Eds.), *Learning to Diagnose with Simulations - Examples from Teacher Education and Medical Education* (chapter 6). Springer.

2020

Kramer, M., Förtsch, C., & Neuhaus, B. J. (2020). Steigern der Unterrichtsqualität – Förderung von Diagnosekompetenzen im Fach Biologie. In S. Habig (Ed.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen: Gesellschaft für Didaktik der Chemie und Physik* (Vol. 40, pp. 210–213). Universität Duisburg-Essen

Tagungsbeiträge

2021

Irmer, M., Frick, D., **Kramer, M.**, Förtsch, C., & Neuhaus, B. J. (2021, March). *Scaffolding in der Lernumgebung DiKoBi*. In U. Harms (Chair), COSIMA meets ProSim et al. Simulationsbasierte Lernumgebungen in der Hochschullehre 22. / 23. März 2021. Digitales Symposium.

2020

Kramer, M., Förtsch, C., & Neuhaus, B. J. (2020, April). *Fostering biology teachers' professional knowledge and effects on professional vision of biology instruction*. Präsentation (Round Table) auf der American Educational Research Association (AERA), San Francisco. (Corona-bedingt abgesagt)

Kramer, M., Förtsch, C., & Neuhaus, B. J. (2020, März). *Einfluss von ePCK für die Diagnose von Unterrichtsqualität im Fach Biologie*. In C. Förtsch (Chair), Förderung und Messung von enacted PCK im Rahmen der universitären Lehramtsausbildung. Symposium auf der 8. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Potsdam. (Corona-bedingt abgesagt)

2019

Kramer, M., Förtsch, C., & Neuhaus, B. J. (2019, September). *DiKoBi – Diagnose von Unterrichtsqualität im Fach Biologie*. In C. Förtsch, B. J. Neuhaus & A. Nehring (Chair), Naturwissenschaftsdidaktische Unterrichtsqualitätsforschung zwischen fachspezifischen und generischen Merkmalen. Symposium auf der Internationalen Tagung der Fachsektion Didaktik der Biologie (FDdB im VBio), Wien.

Kramer, M., Stürmer, J., Förtsch, C., Förtsch, S., & Neuhaus, B. J. (2019, August). *Validating a Simulation-Based Learning Environment Measuring Biology Teachers' Professional Vision*. Präsentation auf der European Association for Research on Learning and Instruction (EARLI), Aachen.

Kramer, M., Stürmer, J., Förtsch, C., Förtsch, S., & Neuhaus, B. J. (2019, März). *Validierung einer simulationsbasierten Lernumgebung zur Erfassung professioneller Unterrichtswahrnehmung mittels Think-Aloud Interviews*. Präsentation auf der 19. Frühjahrsschule der Fachsektion Didaktik der Biologie im VBIO, Bonn.

Kramer, M., Förtsch, C., Förtsch, S., & Neuhaus, B. J. (2019, Februar). *Validierung einer simulationsbasierten Lernumgebung zur Erfassung und Förderung professioneller Unterrichtswahrnehmung und epistemisch-diagnostischer Aktivitäten von angehenden Biologielehrkräften*. In F. Fischer & A. Opitz (Chair), Analyse und Förderung von Diagnosekompetenzen – Erstellung und Validierung von simulationsbasierten Lernumgebungen. Symposium auf der 7. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Köln.

Kramer, M., Stürmer, J., Förtsch, C., Förtsch, S., & Neuhaus, B. J. (2019, Februar). *Teachers' knowledge as prerequisite for the diagnosis of biology instruction*. Poster auf der REASON Winter School 2019, München.

2018

Kramer, M., Förtsch, C., Aufleger, M. & Neuhaus, B. J. (2018, September). *Videoanalyse zum Einsatz digitaler Medien im Biologieunterricht*. Posterpräsentation auf der Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik, Kiel.

München, den 25.03.2021 *Maria Kramer*

8. Appendices

A Task design in DiKoBi *Assess*

The following three tasks (Task *Describe*, Task *Explain*, and Task *Alternative Strategy*) were the same for the six classroom situations embedded in DiKoBi *Assess*. The Figures A1, A2, and A3 show the original German version of the tasks as they were used in the studies.

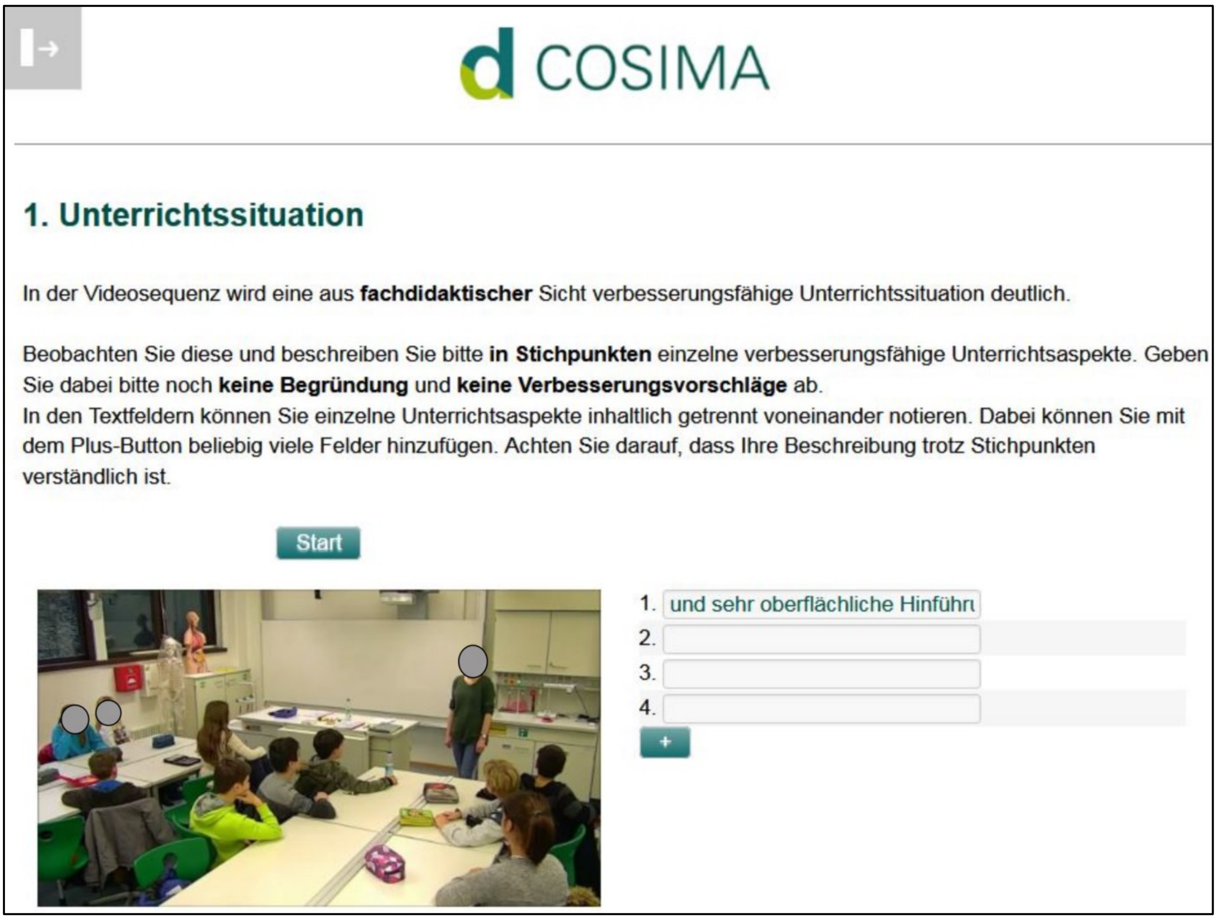

Figure A1. Design of the Task *Describe*. Participants are asked to note challenging teaching aspects they identified in the open text fields on the right (corresponds to the diagnostic activity *generating evidence*).

Figure A2. Design of the Task *Explain*. Participants are asked for didactical rationales (corresponds to the diagnostic activity *evaluating evidence*). Additionally, participants have to estimate their confidence about the answers they have given.

Figure A3. Design of the Task *Alternative Strategy*. Participants are asked for alternative teaching strategies (corresponds to the diagnostic activity *drawing conclusions*). Additionally, participants have to estimate their confidence about the answers they have given.

B Ethical approval by ethics committee

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR PSYCHOLOGIE UND PÄDAGOGIK
MUNICH CENTER OF THE LEARNING SCIENCES
LEARNING SCIENCES RESEARCH METHODOLOGIES

PROF. DR. MORITZ HEENE

Herr Frank Fischer
Im Hause

Prof. Dr. Moritz Heene

Telefon +49 (0)89-2180-72514
Telefax +49 (0)89-2180-99-6888
Mail: heene@psy.lmu.de

Leopoldstr. 13
80802 München

München, 07.12.17

Ihr Antrag an die Ethikkommission

Lieber Herr Fischer,

ich habe mich eingehend mit Ihrer Wiedereinreichung Ihres Antrages für die Sitzung der Ethikkommission am 18.12.2017 bezüglich des Forschungsprojekts

„Analyse und Förderung von Diagnosekompetenzen in simulationsbasierten Lernumgebungen an der Hochschule“

beschäftigt. Gemäß §7, Absatz 4 teile ich Ihnen das Votum der Ethikkommission hiermit mit:

Es bestehen keine Bedenken gegen die Durchführung des Forschungsvorhabens.

Mit freundlichen Grüßen

Prof. Dr. Moritz Heene
Forschungsdekan und Leiter der Ethikkommission

Dienstgebäude
Leopoldstr. 13, Zi. 3520
80802 München

Öffentliche Verkehrsmittel
StadtBus 154, Haltestelle Giselastraße
U-Bahn U3/U6, Haltestelle Giselastraße

Bayerische Landesbank München
Kto. 24 868 BLZ 700 500 00
USt-IdNr. DE 811 205 325

C Guidline for supervisors

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN · COSIMA · Technische Universität München TUM

Informationen zum Ablauf

1. Teil

Der heutige Kurstag ist zweigeteilt. Zuerst erfolgen Wissenstests, im Anschluss schauen Sie sich die Lernumgebung an.

Im ersten Teil werden Sie drei verschiedene Wissenstests bearbeiten. Für jeden Wissenstest ist eine feste Bearbeitungszeit vorgegeben. Deshalb erfolgt die Bearbeitung der Tests nacheinander. Ich teile Ihnen den ersten Test aus und wenn alle den Test vorliegen haben, gebe ich ein Startsignal und starte die Zeit. 5 Minuten vor Ende der Bearbeitungszeit gebe ich Ihnen kurz Bescheid. Sobald die Bearbeitungszeit rum ist, reichen Sie bitte die Bögen an den Rand, sodass ich Sie schnell einsammeln kann.

Dann folgt das gleiche Prozedere für den zweiten Test und im Anschluss auch für den dritten Test.

Sobald ich den dritten Test eingesammelt habe, starten wir mit der Lernumgebung. Kurze Hinweise dazu gebe ich Ihnen dann unmittelbar davor.

Während ich nun die ersten Testbögen austeile, lesen Sie bitte das Informationsblatt vor Ihnen einmal vollständig. Der erwähnte Identifikationscode ist besonders wichtig. Denken Sie bitte immer daran, diesen auf den Tests und in der Lernumgebung anzuwenden, damit die Ergebnisse von Test und Lernumgebung in der Auswertung dann aufeinander bezogen werden können.

→ ***(Versuchsleiter*in teilt PCK-Test aus)***

Starten Sie bitte jetzt mit dem ersten Test. Sie haben dafür 25 Minuten Zeit.

→ ***(5 min vor Zeitende)*** Sie haben noch 5 Minuten Zeit.

(Zeit abgelaufen) Bitte stoppen Sie jetzt Ihre Bearbeitung und geben Sie die Testbögen an den Rand/zu mir.

(Versuchsleiter*in sammelt PCK-Tests ein und verstaut Sie auf einem festgelegten Platz.)

– –

Nun geht es mit dem zweiten Test weiter. Sie haben 15 Minuten Zeit dafür.

→ ***(Versuchsleiter*in teilt CK-Test aus)***

Starten Sie bitte jetzt mit dem zweiten Test.

→ ***(5 min vor Zeitende)*** Sie haben noch 5 Minuten Zeit.

(Zeit abgelaufen) Bitte stoppen Sie jetzt Ihre Bearbeitung und geben Sie die Testbögen an den Rand/zu mir.

(Versuchsleiter*in sammelt CK-Tests ein und verstaut Sie auf einem festgelegten Platz.)

_ _

Nun geht es mit dem dritten Test weiter. Sie haben 10 Minuten Zeit dafür.

→ ***(Versuchsleiter*in teilt PK-Test aus)***

Starten Sie bitte jetzt mit dem dritten Test.

→ ***(5 min vor Zeitende)*** Sie haben noch 5 Minuten Zeit.

(Zeit abgelaufen) Bitte stoppen Sie jetzt Ihre Bearbeitung und geben Sie die Testbögen an den Rand/zu mir.

(Versuchsleiter*in sammelt PK-Tests ein und verstaut Sie auf einem festgelegten Platz.)

2. Teil

Nun folgt der zweite Teil, den Sie an den Computern vor Ihnen durchführen werden.

Sie werden gleich 6 verschiedene Unterrichtssituationen sehen, zu denen jeweils 3 Arbeitsanweisungen folgen. Arbeiten Sie diese bitte zügig durch. Als Orientierung sollten Sie pro Bearbeitung einer Unterrichtssituation 5-8 min anpeilen. Ich werde Ihnen, sobald das erste Unterrichtsvideo startet, jeweils nach 8 Minuten einen zeitlichen Hinweis geben, damit Sie einschätzen können, ob Sie ein gutes Arbeitstempo haben oder etwas zügiger sein sollten. Wenn Sie mit der Bearbeitung des Lernprogrammes durch sind, können Sie leise zusammenpacken und den heutigen Kurstag verlassen.

Seien Sie in …… dann wieder pünktlich …. Uhr in Raum …..

Bitte starten Sie jetzt mit dem Lernprogramm.

(Studierende aktivieren die PCs.)

D Information for participants and declaration of consent

Probandeninformation
DFG Forschergruppe COSIMA

Probandeninformation
Sie sind eingeladen, an unserer Studie zur Diagnosekompetenz von angehenden Biologielehrkräften teilzunehmen.
Bitte lesen Sie die folgenden Informationen sorgfältig durch, bevor Sie zustimmen, an der Studie teilzunehmen. Sollten Sie zusätzliche Fragen haben, können Sie sich jederzeit an die Studienleiterin (maria.kramer@bio.lmu.de) bzw. an die Didaktik der Biologie (Didaktik.Biologie@lrz.uni-muenchen.de) wenden.

Ziel der Untersuchung
Ziel dieser Studie ist es, Kompetenzen der Unterrichtsdiagnose von angehenden Biologielehrkräften anhand von simulationsbasierten Lernprogrammen zu untersuchen. Gegenstand der Diagnose sind somit Unterrichtsprozesse auf Klassenebene und das damit verbundene Lehrerhandeln. Auf langfristige Sicht sollen Förderansätze zur Verbesserung der Diagnosekompetenz für die universitäre Lehrerausbildung entwickelt werden.

Verfahren und Dauer der Untersuchung
Die Aufgabe innerhalb des Lernprogrammes besteht darin, sich Videosequenzen von Biologieunterricht anzusehen und didaktisch problematische Unterrichtssituationen zu diagnostizieren. Dabei werden Ihnen auch Fragen zu Ihrer Person (Alter, Geschlecht, Muttersprache, schulische und universitäre Ausbildung), zu Ihrer Einstellung sowie zur Wahrnehmung des Lernprogrammes gestellt.
Die Bearbeitung des Lernprogrammes dauert insgesamt circa 45 Minuten.

Freiwilligkeit
Die Teilnahme an dieser Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen Ihre Einwilligung zur Teilnahme an dieser Studie widerrufen, ohne dass Ihnen daraus Nachteile entstehen. Ebenso können Sie Ihre Einwilligung zur Speicherung der Daten bis zum Ende der Datenerhebung widerrufen, ohne dass Ihnen daraus Nachteile entstehen.

Datenschutz
Die erhobenen Daten dienen ausschließlich der Erforschung von Möglichkeiten zur Förderung der Diagnosekompetenzen von angehenden Biologielehrkräften im Rahmen der DFG - Forschergruppe "Cosima". Die Daten werden innerhalb der Forschergruppe vertraulich behandelt und vor der Veröffentlichung so aufbereitet, dass ein Rückschluss auf Ihre Person nicht mehr möglich ist.

Verwendung der anonymisierten Daten
Ergebnisse und Daten dieser Studie werden als wissenschaftliche Publikation veröffentlicht. Dies geschieht in anonymisierter Form, d.h. ohne dass die Daten einer spezifischen Person zugeordnet werden können. Die vollständig anonymisierten Daten dieser Studie werden als offene Daten im Internet in einem sicheren, internetbasierten Datenarchiv zugänglich gemacht. Damit folgt diese Studie den Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) und der Deutschen Gesellschaft für Psychologie (DGPs) zur Qualitätssicherung in der Forschung. Nicht vollständig anonymisierbare Daten (wie etwa Audio- und Videoaufzeichnungen) werden nicht veröffentlicht.

○ Hiermit versichere ich, dass ich die oben beschriebenen Probandeninformationen verstanden habe und mit den genannten Teilnahmebedingungen einverstanden bin.

Hinweise zur Verwendung der Videos im Lernprogramm

Wie Sie schon erfahren haben, werden Sie im Lernprogramm DiKoBi mit Unterrichtsvideos arbeiten. In diesem Zusammenhang bitten wir Sie, die folgenden Informationen zur Verwendung der Videos im Lernprogramm sorgfältig durchzulesen, bevor Sie diesen zustimmen.

Ich verpflichte mich, die Videos nicht aus dem Lernprogramm zu kopieren, sie nur für den mir ausgehändigten Zweck zu nutzen und sie nur unter den Fragestellungen zu bearbeiten, welche im Lernprogramm genannt werden.

Versehentliche Abweichungen von der oben beschriebenen Verwendung der Videos im Lernprogramm teile ich der Abteilung für die Didaktik der Biologie umgehend mit.

○ Ich habe die Informationen zur Verwendung der Videos im Lernprogramm DiKoBi gelesen, verstanden und erkläre mich damit einverstanden.

E Electronic supplemental material

The dissertation further contains an electronic appendix consisting of the following contents:

- Study 1: Material (information texts) used for the intervention
  - Materials on PCK
  - Materials on CK
  - Materials on PK
  - Materials on combination of PCK/CK/PK
- Study 2: Material (lectures notes) used for the intervention
  - Lecture notes for separated instruction
  - Lecture notes for integrated instruction

All material is confidential. Please do not pass on the material.

Final note:
The coding manual created and used for the analyses was published as a technical report in "MCLS reports", an edited series of text published in the context of the Munich Center of the Lerning Sciences (MCLS). The coding manual is available under:

Kramer, M., Förtsch, C., Stürmer, J., & Neuhaus, Birgit J. (29. November 2021). DiKoBi: Kodiermanual DiKoBi - Messung von Diagnosekompetenzen von Biologielehrkräften im Biologieunterricht mit einem videobasierten Simulationstool. Munich Center of the Learning Sciences: MCLS Reports, Nr. 2. https://doi.org/10.5282/ubm/epub.77972