International Journal of Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tsed20

Scaffolding pre-service biology teachers' diagnostic competences in a video-based Learning environment: measuring the effect of different types of scaffolds

Marie Irmer, Dagmar Traub, Maria Kramer, Christian Förtsch & Birgit J. Neuhaus

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
https://doi.org/10.1080/09500693.2022.2083253

Scaffolding pre-service biology teachers' diagnostic competences in a video-based Learning environment: measuring the effect of different types of scaffolds

Marie Irmer, Dagmar Traub, Maria Kramer, Christian Förtsch and Birgit J. Neuhaus

Biology Education, Faculty of Biology, Ludwig-Maximilians-Universität München, Munich, Germany

ABSTRACT

Diagnosing is an important aspect of teachers' professional competence. To successfully diagnose classroom situations, one has to apply pedagogical content knowledge (PCK) in diagnostic activities (DA). Simulated classroom situations with embedded scaffolding can be an effective way of practising skills needed to foster diagnostic competences. In an intervention study with pre-post-test design using the video-based learning environment *DiKoBi*, we investigated the effects of scaffolds on the diagnostic competences of 57 pre-service biology-teachers in an early stage of university education. For the intervention, we developed two types of scaffolds: one providing additional PCK (PCK-scaffold) and one focusing on the diagnostic activities (DA-scaffold). Pre-service teachers were divided into four groups depending on the type of scaffold they received during the intervention (PCK-, PCK + DA-, DA- and control group). Pre-service teachers receiving PCK-scaffolds showed statistically significantly higher diagnostic competences in the post-test than in the pre-test. This includes the PCK- as well as the PCK + DA-group. For the other two groups, no significant improvement can be reported. Thus, we assume that the diagnostic competences of pre-service biology teachers can effectively be fostered by including PCK-scaffolds in *DiKoBi* and hence a minimum of PCK is necessary for successfully diagnosing in simulated classroom situations.

ARTICLE HISTORY
Received 8 September 2021
Accepted 24 May 2022

KEYWORDS
Video-based learning environment; scaffolding; diagnosing

1. Introduction

Teachers in a classroom setting are constantly required to deal with diagnostic situations whenever they interact with their students. This includes diagnosing students' misconceptions or a current classroom situation (e.g. teachers' planning of the lesson, students' behaviour, classroom management). To succeed in such diagnostic situations, diagnostic competence is necessary; hence approaching this diagnostic competence is necessary in university pre-service teacher education. Diagnostic competences contain two components: The knowledge in the specific field and the process a learner has to master while diagnosing (Heitzmann et al., 2019). Diagnosing in educational contexts can

CONTACT Marie Irmer marie.irmer@bio.lmu.de Biology Education, Faculty of Biology, Winzererstrasse 45/II, Ludwig-Maximilians-Universität München, Munich, Bavaria 80797, Germany

focus on different aspects such as the assessment of students' achievements, the examination of students' characteristics relevant for performance and learning as well as on the analysis of classroom situations, the impact of instruction and contextual factors (Glogger-Frey et al., 2018). One aim of diagnosing in classroom situations is the establishment of learning conditions that enable learners to achieve subject-specific and/or pedagogical learning goals. Research has shown that instructional quality is positively related to student outcomes (Seidel & Shavelson, 2007). If pre-service teachers learn about instructional quality by diagnosing classroom situations, we assume that this will then be positively related to their students' outcomes in the future. Overall, diagnosing in classroom situations can be quite a complex task especially for learners in an early stage of expertise development (Chernikova et al., 2020). To reduce complexity, simulation-based learning offers learning close to real-life practice while still allowing to adjust a certain situation to a learners' needs (Chernikova et al., 2020). One possibility to help learners with achieving their learning goals, is to offer additional support in a learning situation. When scaffolds are added to learning environments, learners can reach a new level of learning (Collins et al., 1988; Plass & Pawar, 2020). There are many different ways to design a scaffold depending on the learner needs, e.g. providing additional information or supporting the process of a certain task. This study aims to compare different types of scaffolds implemented in a simulation-based learning environment.

2. Theoretical background

2.1. Professional competence of a teacher in the context of diagnosis

Professional competence as described by Blömeke et al. (2015) is considered as a continuum between dispositions (cognition and affect-motivation) and performance (observable behaviour) mediated by situation-specific skills (perception, interpretation and decision making). In the context of diagnosing instructional quality, it has been described as diagnostic competences (Heitzmann, 2014; Ohle & McElvany, 2015). A competent teacher, creating a high-quality learning environment in the classroom, is so continuously required to deal with diagnostic situations. Considering the model of Blömeke et al. (2015), every diagnosis made by applicating situation-specific skills, is influenced by the professional knowledge as well as by the affective-motivation of a teacher. A teacher's professional knowledge can be categorised into three knowledge facets: pedagogical-psychological knowledge (PK), content knowledge (CK) and pedagogical content-knowledge (PCK) (Baumert & Kunter, 2013; Shulman, 1986). Interdisciplinary knowledge about classroom management, learning strategies and teaching methods can be assigned to a teacher's PK. CK includes subject-specific knowledge, e.g. knowledge about biological facts and concepts. To make these concepts accessible for students, a teacher needs PCK, which includes knowledge about students' misconceptions, teaching strategies and subject-specific structures of instruction (Schmelzing et al., 2013; Shulman, 1986; Voss & Kunter, 2013). These three knowledge facets influence the instructional quality and thus students' learning outcomes (Förtsch et al., 2016; Schmelzing et al., 2013; Voss & Kunter, 2013).

2.2. Diagnostic activities

According to Heitzmann et al. (2019), diagnostic competences are shown when knowledge in diagnostic activities is applied. The goal of applying diagnostic activities is the collection and interpretation of data in order to make high-quality decisions (Heitzmann et al., 2019). To describe the skills needed to perform diagnostic competences using professional knowledge, Fischer et al. (2014) elaborated the concept of *diagnostic activities*. The concept of diagnostic activities describes eight activities relevant for generating knowledge in the specific domain, e.g. teaching (Fischer et al., 2014; Heitzmann et al., 2019). Kramer et al. (2021) assume four of these eight activities to be relevant to diagnose in classroom situations - *identifying problems*, *generating evidence*, *evaluating evidence* and *drawing conclusions* - as these four activities represent steps during the assessment of a diagnosis. A teacher's ability to identify a challenging event within a teaching situation refers to the diagnostic activity of *identifying problems*. Teachers *generate evidence* when describing challenging situations without justification or evaluation, they then *evaluate* these evidences by linking their observation to theoretical concepts and explaining on a professional knowledge base. *Drawing conclusions* means to improve a challenging situation by proposing concrete alternatives for action (Fischer et al., 2014).

The construct of professional vision has become popular to operationalise situation-specific skills in teacher education (Gaudin & Chaliès, 2015; Seidel & Stürmer, 2014). It comprises two components for professionally observing and interpreting classroom situations: noticing and reasoning (Borko, 2004; Seidel & Stürmer, 2014). Noticing is described as concentrating on events relevant for students' learning whereas reasoning is subdivided into the three aspects *description*, *explanation* and *prediction* (Seidel & Stürmer, 2014). Seidel and Stürmer (2014) define *description* as focusing on but not judging relevant aspects in a teaching situation. *Explanation* means using subject-specific theories and concepts to reason about the described aspect, and *prediction* refers to consequences derived from the observed aspect, that was described and explained before (Seidel & Stürmer, 2014).

Both concepts, *diagnostic activities* and *professional vision*, specify situation-specific skills in diagnostic situations. Kramer et al. (2021) stated that these two concepts largely overlap, but also complement each other to some extent. For example, the diagnostic activity *evidence generation* fully corresponds to the professional vision construct *description*. In the context of diagnosing classroom situations, the concept of diagnostic activities is preferably used.

2.3. Video-based tools in teacher education

The demand for practical orientation in teacher education is persistent (Blömeke et al., 2015). Using video-based tools showing authentic classroom situations is an effective method to guarantee more practical insights for pre-service teachers during teacher education (Südkamp et al., 2008). Video-based tools reduce the complexity of reality and can provide additional support (e.g. in form of scaffolds), and thus, they can help to ensure that pre-service teachers are not overburdened when they work on practice-oriented tasks (Gartmeier et al., 2015; Grossmann et al., 2009). Furthermore, video-based instruments are efficiently used to measure a teachers' professional vision and as a tool for professional development (Roth et al., 2017; Seidel & Stürmer, 2014). Videos can be included in teacher education to practice reasoning about instructional quality and can improve these skills as well as the professional knowledge of pre-service teachers (Behling et al., 2019; Gaudin & Chaliès, 2015; Santagata & Guarino, 2011).

2.4. Scaffolding

Scaffolding is often associated with the *Zone of Proximal Development* where a learner is able to complete challenging tasks with external support, that could not be completed successfully without support yet (Vygotskij, 1979). Wood et al. (1976) initially introduced the term *scaffold* as a provided help in a learning setting to assist the learner achieve a goal that could not be achieved without this assistance. They defined six functions for scaffolding: (1) creation of situational interest, (2) reduction of complexity of the task, (3) keeping the learners focused, (4) marking of critical features of the task, (5) controlling possible frustration of the learner and (6) demonstration of solutions for the task (Wood et al., 1976). Scaffolds offering additional content information can reduce the complexity of a task and can keep the learner focused. A scaffold marking critical features of a task could be a guidance on how the task has to be performed. We define a scaffold as the presented help to the learner, whereas scaffolding means the process of helping a learner to achieve a certain learning goal by including scaffolds. Hannafin et al. (1999) distinguished between conceptual, metacognitive, procedural, and strategic scaffolding. There are related methods and mechanisms for every type of scaffold, including Vygotskian scaffolding such as the provision of explicit hints and prompts for conceptual scaffolding (Hannafin et al., 1999). Scaffolding can be subdivided into *context-specific* and *generic scaffolding* (Belland et al., 2013). Context-specific scaffolds can, for example, be prompts containing specific content to the problem addressed in a learning situation (Belland, 2017). Generic scaffolds are not tailored especially for one context and can, for example, structure a task by giving additional guidance in how to complete it (Belland, 2017). The goal of scaffolding is not only to enable the learner to complete a tasks in the moment when the scaffold is presented, but also to improve one's performance in future tasks (Reiser, 2004).

Scaffolding can be a highly effective intervention for increasing the learner's abilities to complete difficult tasks (Belland et al., 2017; Quintana et al., 2004), and thus reach the Zone of proximal development. Meta-analyses showed that scaffolding providing a high level of guidance through a learning environment is very effective for leaners with little prior knowledge (Chernikova et al., 2020). When scaffolds are implemented in a learning environment that requires problem solving, researchers found significant improvement for learners receiving scaffolds (Cho & Jonassen, 2002; Simons & Klein, 2007). Scaffolding can so help learners to complete tasks, to integrate information and to solve problems (Simons & Klein, 2007). Scaffolding can also be an effective support in diagnosing classroom situations in video-based learning environments. For example, according to Belland et al.'s (2013) classification, *context-specific scaffolds* that provide additional information (e.g. PCK) can be used to help learners to apply their context-specific knowledge in a classroom situation. This additionally provided declarative knowledge, also described as 'knowing that' (Förtsch et al., 2018), is all knowledge about terms, principles, and facts in a specific domain. *Generic scaffolding* provides the opportunity to address the diagnostic activities. Pre-service teachers are given additional information on how to apply the diagnostic activities, such as generating evidence before evaluating evidence, as they watch the classroom videos. According to Chernikova et al. (2020), learners with higher prior knowledge benefit most from scaffolds providing additional action-related information to support the process of diagnosis.

3. Research questions and objectives

Supporting pre-service teachers to diagnose in classroom situations, scaffolding could be an effective intervention. To investigate what kind of scaffold helps to foster diagnostic competences in a video-based learning environment, according to Belland et al.'s (2017) classification, two different types of scaffolds were developed: *context-specific scaffolds* that present declarative knowledge (content-related PCK-scaffold), and *generic scaffolds* that address the diagnostic activities (DA-scaffolds). These two scaffolds refer to the two components of diagnostic competences (Heitzmann et al., 2019). In this case, the two components would be the PCK and three diagnostic activities.

The main goal of the present study is to find out what type of scaffold is more effective in fostering pre-service teachers' diagnostic competences.

We derived the following research questions:

> RQ1: Can the use of different scaffolds in the video-based learning environment DiKoBi improve the diagnostic competences of pre-service biology teachers?

> RQ2: How do different types of scaffolds differ in their effects on pre-service biology teachers' diagnostic competences?

4. Methods

4.1. Sample

The sample consisted of 57 pre-service biology teachers for German primary and secondary schools (14.0% male, 84.2% female, 1.8% without specification). The average age was 22.3 years ($SD = 3.32$, Min = 19, Max = 35). All pre-service teachers were in an early stage of their university education (semester: $M = 2.24$, $SD = 0.79$, Min = 2, Max = 6).

4.2. The video-based learning environment DiKoBi

All studies were conducted using the video-based learning environment *DiKoBi* (German acronym for diagnostic competences of biology teachers in biology classrooms) (Kramer et al., 2020).

Within *DiKoBi*, videos of five different challenging biology classroom situations are shown. Each challenging classroom situation focuses on another dimension of biology-specific instructional quality (Kramer et al., 2020) (see Figure 1). Three separate versions of *DiKoBi* show three consecutive biology lessons on the topic *skin*, where all five challenging biology classroom situations are specified for another subtopic of the main topic *skin*: *DiKoBi^{I}*: *Skin as a sensory organ*, *DiKoBi^{II}*: *Protective functions of the skin* and *DiKoBi^{III}*: *Regulation of body temperature*. Additionally, *DiKoBi* can be used as a

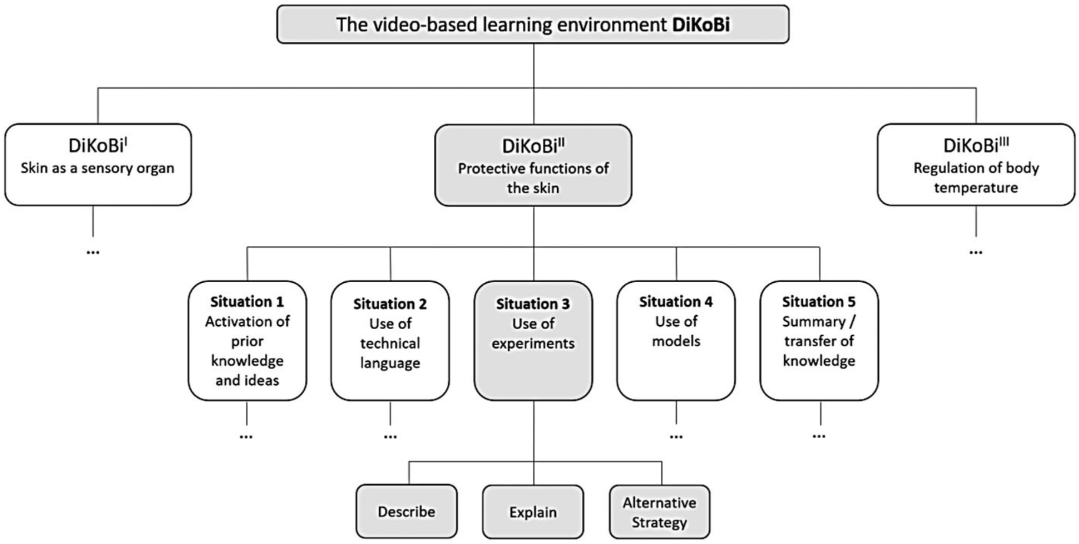

Figure 1. The learning environment DiKoBi. Each lesson contains five challenging classroom situations focussing on five different aspects of biology-specific instructional quality. For each situation, the pre-service teachers have to complete three tasks (describe, explain and alternative strategy). The concept is shown for one example (marked in grey) (Kramer et al., 2020).

measuring instrument without scaffolds (*DiKoBi Assess*) and a learning environment in which different types of scaffolds are embedded (*DiKoBi Learn*).

While working on the video-based learning environment, the pre-service teachers take over the role of an observer focusing on the teacher's instruction in a classroom situation, which was perceived as authentic (Kramer et al., 2020). In the shown classroom situations in *DiKoBi*, relevant features of the focused dimension are missing, which the pre-service teachers have to identify while watching the videos. For this, they have to complete the three tasks *describe*, *explain* and *alternative strategies* by taking notes in open text fields (cf. diagnostic activities, Fischer et al., 2014). In the task *describe*, the pre-service teachers are required to describe critical aspects value-free and without justifications, this task measures the diagnostic activities *identifying problems* and *generating evidence*. In the second task (*explain*), the described observation should be linked to didactical theories, concepts and technical terms to justify why it is considered as problematic. This task refers to the diagnostic activity *evaluating evidence*. The diagnostic activity *drawing conclusions* is represented in the task *alternative strategy*. The pre-service teachers have to describe alternative courses of action, giving detailed descriptions of examples linked to theories and concepts. For an example, see Table 1.

Within this study, three versions of *DiKoBi* were used: *DiKoBi^{I} Assess*, *DiKoBi^{II} Learn* and *DiKoBi^{III} Assess*. The dimensions of biology-specific instructional quality are the same for all three lessons; just the subtopic differs.

4.4. Design of the scaffolds

According to the classification of Belland et al. (2013), two different types of scaffolds were developed to be included in the learning environment *DiKoBi^{II} Learn*: (I) content-related PCK-scaffolds, which correspond to the context-specific scaffolds, (II) scaffolds focussing on the diagnostic activities (DA-Scaffolds), which correspond

Table 1. Example of a diagnosis in the learning environment.

Situation	Teacher action	Task *describe*	Task *explain*	Task *alternative strategy*
Beginning of the lesson, cognitive activation, interest and motivation	Lesson starts with the teacher asking what the students remember from the last lesson (sensory organs) and then the teacher introduces the topic of the current lesson	Students repeat prior knowledge about the sensory organs but do not deal with the subject of the previous lesson in more depth Teachers names the topic of the lesson instead of letting the students formulate it Questions asked by the teacher are reproductive, no explanations are required No problem orientation	There is no cognitive activation of the students. The students should connect their prior knowledge to subordinated concepts	The teacher could start the lesson with a cognitive conflict leading to a problem-orientated question as the topic of a lesson. Instead of just letting the students repeat what they remember from the last lesson, the teacher could activate their prior knowledge by asking question that require concepts to answer
		Missing motivation (no example, no context, nothing new or surprising, no relevance for everyday life)	No sufficient activation of the students. No catch-component for creating situational interest	An experiment, something surprising, bringing up more relevance for everyday life could serve as a catch-component to create situational interest

to the generic scaffolds. The scaffolds were designed based on the two components of diagnostic competences; meaning the knowledge in a specific field (here: PCK) and the process (here: applying diagnostic activities). The content-related PCK-scaffolds focused on the dimensions of biology-specific instructional quality shown in the classroom situations. They are intended to help pre-service teachers with low prior pedagogical content knowledge to focus on the aspect asked for by giving further information helping with the content to complete the tasks (e.g. see Figure 3, Wood et al., 1976). The scaffolds were developed based on relevant literature on subject-specific dimensions of instructional quality and corresponding features, as well as literature on specific skills relevant to succeed in diagnostic processes (see Table 2). We decided to use scaffolds providing additional information to support a pre-service teachers' PCK, because this knowledge facet is the determining factor for effective biology-teaching (Förtsch et al., 2016) and thus can help leaners to complete the tasks in the learning environment. Additionally, all aspects of the scaffolds are in accordance with theoretical aspects that were listed in a coding manual and used for coding pre-service teachers' answers from the open text fields in *DiKoBi* (Kramer et al., 2020). We decided to use rather short texts of approximately the same length for all five situations (number of words: between 159 and 168). Each scaffold included keywords that represent important science-specific terms. Concepts or principles with regard to literature were bold printed (see Figure 3).

The second type of scaffolds (DA-scaffolds) focuses on the diagnostic activities *identifying problems* and *generating evidences* (prompted in the task *describe*), *evaluating*

Table 2. Overview for PCK-scaffolds with the associated dimension of instructional quality, keywords, literature and length of the scaffolds.

Dimension of instructional quality	Keywords	literature (examples)	length
Cognitive activation and creation of situational interest	Activation, cognitive, prior knowledge, student ideas, conflict, problem/focus question, catch, situational, interest, interesting, alienated, everyday life, competition, motivation, surprise, curiosity, closeness to the self, affective-motivational, motivating, problem(orientated)	Dorfner et al. (2019); Förtsch et al. (2016); Schmiemann et al. (2011)	159 words
Use of technical language	Communication, technical language, everyday language, education language, technical terms, manageable, number, explanation/ explain, science level, educational level, competence, passive and subordinate clause constructions, language sensitivity, context, linking, abstraction level, retrieve	Riebling (2013); Nodari and Steinmann (2008)	168 words
Use of experiments	Ask a question, hypothesis, plan, conduction, evaluation, analysis, interpretation, return to something, knowledge acquisition, scientific method, question, data, process related, phase	Mayer et al. (2008); Arnold et al. (2014)	168 words
Use of models	Visual aids, working method, thinking method, knowledge acquisition, model criticism, matches, boundaries, abbreviations, accessories, excerpt, relevant, misconception, hypothesis, scientific method, comparison, original, colour choice, deal	Upmeier zu Belzen and Krüger (2010); Nowak et al. (2013); Oh and Oh (2011)	165 words
Summary/ transfer of knowledge	Back reference, continuous thread, learning progress, application-related, checking, cognitive, activation, reproduction, linking, concept-orientated, transfer, problem-orientated, context, summary, linking degree, structure and function	Dorfner et al. (2019); Wadouh et al. (2014); Meisert (2012);	165 words

evidences (prompted in task the *explain*) and *drawing conclusions* (prompted in the task *alternative strategy*) (Jahn et al., 2014). This scaffold contains no content-related information on biology-specific instructional quality; instead, it contains additional support on how to perform the required diagnostic activity (see Table 3).

4.4. Study design

The study is part of the DFG project *COSIMA* (German acronym for: Facilitating diagnostic competences in simulation-based learning environments in the university context). Research within *COSIMA* focuses on how simulation-based learning environments in a university context can be designed and implemented to facilitate the acquisition of diagnostic competences, in particular in early and intermediate stages of competence.

The current study was conducted as an intervention study with pre–post-test-design (see Figure 2). In the pre-test, the pre-service teachers completed the video-based learning environment *DiKoBi*^{I} *Assess*, measuring diagnostic competences. One week later, within the intervention, the pre-service teachers went through *DiKoBi*^{II} *Learn*. During this intervention, the pre-service teachers were divided in four treatments. For each treatment, different types of scaffolds were included in *DiKoBi*^{II} *Learn*. In the first treatment, the pre-service teachers received content-related PCK-scaffolds (PCK-group), in the second

Table 3. DA-scaffolds in DiKoBi.

Task in *DiKoBi*	Corresponding diagnostic activity	Scaffold
Describe	Identifying problems, generating evidence	*Identifying and describing problems* Use the text field to describe a problem that you have identified. Describing, it is important to present the situation as *value-free* as possible and *without justifications*. Focus purely on features and *details of the situation* or teacher behaviour that you consider critical
Explain	Evaluating evidence	*Analysing and justifying* Highlight the relevance of your identified problems by using your professional knowledge to justify why each observation is problematic. To do this, *link* your observation with *didactical theories*, *concepts* and *technical terms* that you name and explain in your justification
Alternative strategy	Drawing conclusions	*Drawing consequences from the described problems* Based on the aspects you have criticised or identified as problematic, describe *alternative courses of action* for the teaching situation. Be *specific* with your alternative and give *examples and corresponding reasons* that speak for your chosen alternative(s) in a comprehensible and *detailed* way.

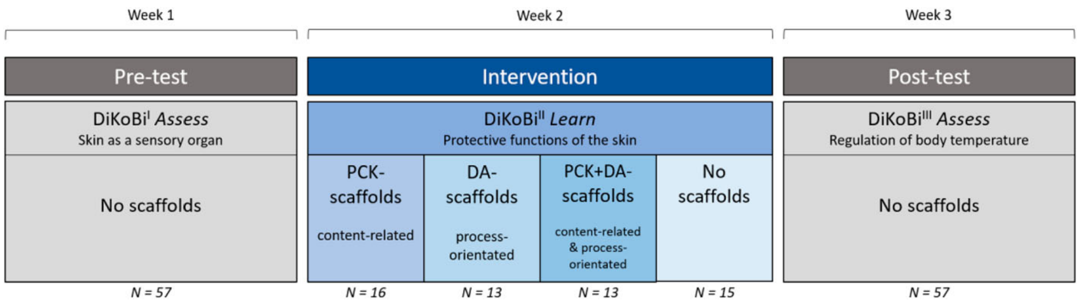

Figure 2. Study design.

treatment they received scaffolds focusing on the diagnostic activities (DA-group), the third treatment contained both scaffolds on PCK and diagnostic activities (PCK + DA-group) and the fourth treatment did not include any scaffolds (control group). The content-related PCK-scaffold is presented during the video, the scaffold focusing on a diagnostic activity pops up before answering the task. Both types of scaffolds can only be read once in the learning environment. They disappear after the pre-service teacher clicks on 'next' and cannot be re-read later in the task. In the following week, the post-test took place. All pre-service teachers completed the learning environment *DiKoBi^{III} Assess*, measuring again the diagnostic competences. At any time of measurement, we used the video-based learning environment *DiKoBi* for measuring the quality of diagnostic activities (*evidence generation, evidence evaluation* and *drawing conclusions*) as well as the accuracy of the diagnosis.

4.5. Measurements and data analysis

4.5.1. Using the number of keywords used as a measure for diagnostic competences

To investigate whether pre-service teachers show better diagnostic competences because they use scaffolds in a way of including content presented in the scaffolds for completing the tasks, we decided to identify keywords that can then be found in the pre-service

When the teacher introduces a new topic in class, the **activation** of students, especially **cognitively**, is important. This should involve the **activation of** the students' **prior knowledge**, **exploring students' ideas** and, if appropriate, generating a **cognitive conflict**. For example, alienated illustrations, which trigger doubt, confusion or ambiguity, can be used and thus have a cognitively activating effect. Ideally, the introduction should lead to the topic of the lesson, be present for the students in form of a **problem or focus question** and should be referred to again at the end of the lesson.

The use of problems relevant to everyday life, alienated illustrations or even the conception of the lesson as a competition can also arouse the learners' motivation for the subject matter. Everyday examples as well as the use of appropriate pictures serve to generate emotions such as surprise or curiosity, but also closeness to the self in order to catch the learners affective-motivationally. These approaches and methods are also known as **catch-components**, which should be present when introducing a new topic in order to arouse **situational interest**.

Figure 3. Content-related scaffold for the first classroom situation. Bold-printed keywords are marked in yellow, keywords we decided for on top of the bold-printed ones are marked in blue. In the original scaffold, no words are marked.

teachers' solutions (see Figure 3). These keywords refer to the relevant dimensions of biology-specific instructional quality that need to be included for a good diagnosis. Thus, by analysing the use of keywords, we take the quality of pre-service teachers diagnoses into account.

To complete tasks in the learning environment, the pre-service teachers needed to write their answers in open text fields. We decided to count the number of keywords used in the participants' answers (in every of the three tasks, for every situation) to measure the effectiveness of a scaffold. The keywords are specific for the aspect of biology-specific instructional quality, so that one could assume that participants using these words to complete the tasks in *DiKoBi* would achieve a higher score. Thus, the more keywords were used, the more effective the scaffold is considered.

The coding of pre-service teachers' answers in *DiKoBi* was done according to Kramer et al. (2021) by two independent and trained coders ($\kappa = 0.93$). This coding led to a sum score for every challenging situation. For every open text field, a pre-service teacher could get up to two (or three for the description task) points (see Table 4). To get the maximum of points, pre-service teachers' answers did not only have to refer to the correct subject-specific content, but the diagnostic activity, e.g. *generating evidence*, had to be performed adequately as well.

For example, two points for *generating evidence* (task *describe*) were given in case pre-service teachers described an identified event detailed and comprehensible by referring to specific details visible in the video (Kramer et al., 2021). The sum score was calculated for every pre-service teacher by adding up the points for all answers from the open text fields and represents a measure of pre-service teachers' diagnostic competence. As scaffolding is intended to support the diagnostic process (Chernikova et al., 2020), we assumed that the number of keywords used in the open text fields correlated with the sum score for each situation. Therefore, we first calculated the sum score of pre-service teachers' overall competences to diagnose the classroom situations of the pre-test and the

Table 4. Overview of the quality levels and their operationalisations for the assessment of diagnostic activities (Kramer et al., 2021).

Task	Code		Description
All tasks	−99	Missing	No answer was given
	0	Not accurate	The given answer (*description*, *explanation*, *alternative strategy*) is not accurate
Describe	1	Unsystematic, incomplete description	Non-specific, superficial description of the identified event; weak/no focus on specific details
	2	Systematic, complete description	Several details of the identified event are described with attributes; a strong focus on specific details (visible in the video)
Explain	1	Empty phrase	The statement is more of an everyday phrase than an explanation, partly meaningless
	2	Simple reference to concepts/ theories	Appropriate to or based on the corresponding description, the subject-specific pedagogical theory is named as a keyword or embedded as a phrase in a sentence
	3	Comprehensive explanation	Observation and theory are related to each other.
Alternative strategy	1	Just one alternative (described non-specific)	One alternative strategy is described. The description is rather general. Concrete examples/reference are missing
	2	Several alternatives (described in detail and with examples)	Several appropriate suggestions for improvement are described with concrete examples and references; it is explained to what extent the alternative strategy improves what has been criticised before

overall number of keywords used, and then we examined their correlation. A strong positive correlation could be revealed ($r = .734$, $p < .001$). Thus, the number of keywords used can be considered a valid measure for the diagnostic competences of the pre-service teachers. We counted the number of keywords each pre-service teacher used separately for the pre-test (*DiKoBi^I Assess*), intervention (*DiKoBi^II Learn*) and post-test (*DiKoBi^III Assess*). This led to three scores (*number of keywords used for pre-test, intervention and post-test*) for each pre-service teacher.

4.5.2. Data analysis

Coded data were analysed using the statistics package *IBM SPSS Statistics 2*6. All calculations were made using these total scores.

RQ1. To test if the learning environment DiKoBi with scaffolds implemented can improve the diagnostic competences of the pre-service teachers, first, a mixed ANOVA (Analysis of variances) was performed. As assessed by Box's test ($p = .031$), there was no homogeneity of covariances, so that a repeated measures analysis of variance (RM-ANOVA) was performed for each group separately (Field, 2009). A Greenhouse-Geissler-adjustment was applied to correct for violations of sphericity.

RQ2. To test significant differences between the four groups during the pre- and the post-test, a one-way ANOVA was performed for the pre-test and an ANCOVA (Analysis of covariances) with the pre-tests' number of keywords used as a covariate for the post-test. Due to a significant Levene-Tests ($p > .001$), a Kruskal-Wallis-test was used to calculate differences between the groups in the intervention as suggested by Field (2009).

5. Results

5.1. RQ 1

In the following, results referring to RQ1 (improvement of diagnostic competences through scaffolds) are reported.

5.1.1. PCK-group

A repeated measures ANOVA with a Greenhouse-Geisser adjustment showed a statistically significant difference in the number of keywords used between the pre-test and the post-test with the pre-service teachers scoring higher in the post-test ($F(1.00, 15.00) = 12.16$, $p = .003$, *partial* $\eta^2 = .448$, $N = 16$). The Greenhouse-Geisser adjustment was used to correct for violations of sphericity. For a descriptive overview, see Table 5.

5.1.2. PCK + DA-group

Due to violations of sphericity, a Greenhouse-Geisser adjustment of the repeated measures ANOVA was made. Results showed that the number of keywords used was statistically significantly higher in the post-test than in the pre-test ($F(1.00, 20.95) = 12.16$, $p = .001$, *partial* $\eta^2 = .636$, $N = 13$). For a descriptive overview, see Table 5.

5.1.3. DA-group and control-group

For the DA-group ($F(1, 12) = .244$, $p = .630$, $N = 13$) and for the control-group ($F(1, 14) = 2.74$, $p = .12$, $N = 15$), a repeated measures ANOVA showed no significant differences between the pre-test and the post-test (for descriptive statistics, see Table 4).

5.2. RQ 2

The following results refer to the differences of the number of keywords used between the four groups for the pre-test, the intervention and the post-test. The descriptive statistics are shown in Table 5. For a comparison of the four groups during the three times of measurement, see Figure 4.

Table 5. Descriptive statistics for pre-test, intervention and post-test for each group, including number of pre-service teachers, mean and standard deviation for the number of keywords used.

Time of measurement	Group	*N*	*M*	*SD*
Pre-test	PCK	16	5.25	6.12
	PCK + DA	13	4.69	4.40
	DA	13	4.15	3.11
	Control	15	2.67	2.61
Intervention	PCK	16	25.13	15.11
	PCK + DA	13	16.77	15.34
	DA	13	7.15	4.24
	Control	15	5.00	3.38
Post-test	PCK	16	9.37	5.05
	PCK + DA	13	8.15	4.67
	DA	13	4.69	3.07
	Control	15	3.93	2.60

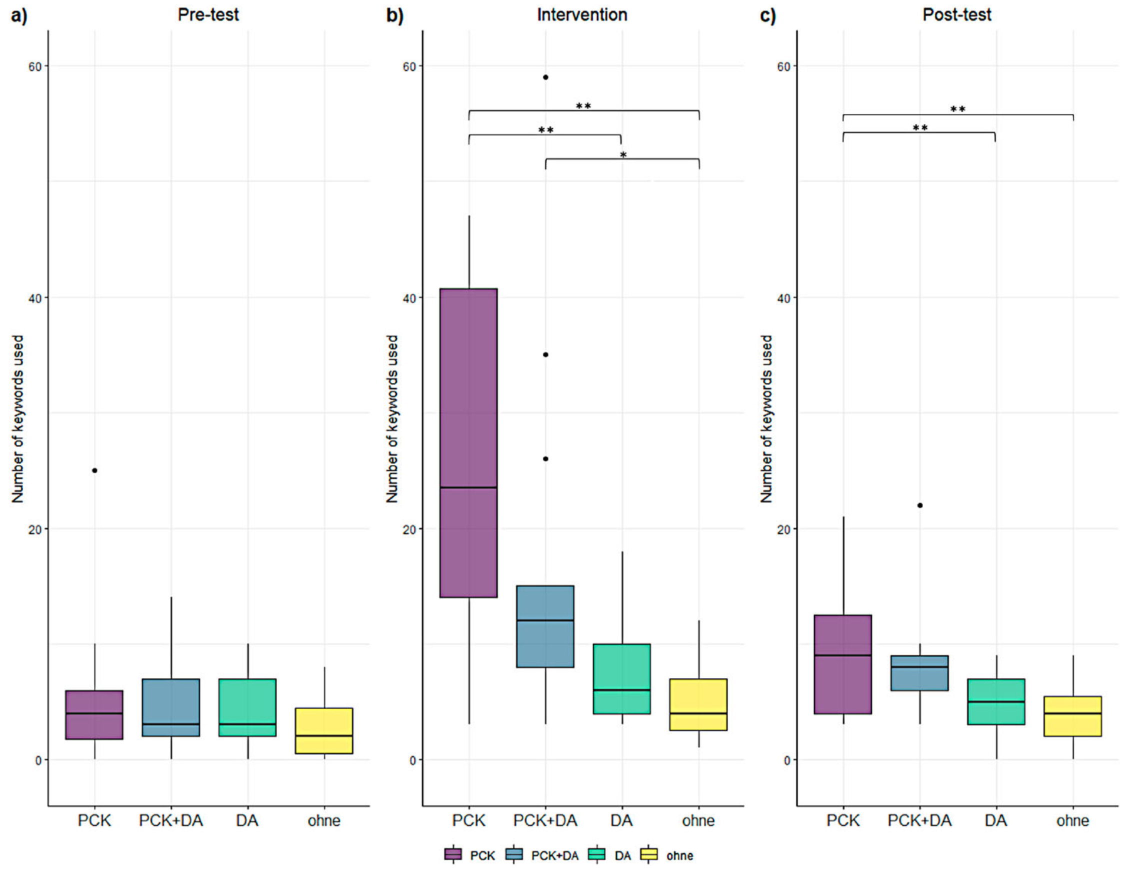

Figure 4. Boxplots showing the four groups during all three times of measurement in comparison.

5.2.1. The pre-test

An ANOVA showed that there are no significant differences of number of keywords used between the four groups ($F(3, 53) = 0.986$, $p = .406$, $N = 57$). Thus, the pre-service teachers did not differ in their diagnostic competences at the pre-test.

5.2.2. The intervention

Due to a significant Levene-test ($p < .001$), a Kruskal-Wallis-test was used to calculate differences between the groups. This test showed that in the four groups the number of keywords used differed (*Kruskal–Wallis* $H = 24{,}444$, $p < .001$, $n = 57$). Furthermore, a post-hoc test with Dunn-Bonferroni adjustment revealed that there are significant differences between the PCK- and the control-group ($z = 4.54$, $p < .001$, $r = .81$), the PCK- and the DA-group ($z = 3.27$, $p = .006$, $r = .61$), as well as between the PCK + DA and the control group ($z = 3.11$, $p = .011$, $r = .62$). The largest difference is found between the PCK- and the control-group (20.13 keywords in mean), followed by the difference between the PCK- and the DA-group (17.98 keywords) and the PCK + DA- and the control-group (11.77 keywords in mean).

5.2.3. The post-test

An ANCOVA showed the type of scaffold presented in the intervention has a significant influence on the number of keywords used in the post-test ($F(4, 52) = 15.05$, $p < .001$, partial $\eta^2 = .537$, $N = 57$). The effect strength according to Cohen is $f = 1.08$, thus showing a strong effect. Post-hoc test with Bonferroni-adjustment shows that there are statistically significant differences between the PCK group and the DA group ($p = .008$) as well as between the PCK group and the control group ($p = .008$) (see Table 5).

6. Discussion and conclusion

The aim of this study was to investigate the effects on the diagnostic competences of different types of scaffolds embedded in the learning environment *DiKoBi*.

6.1. The effects of the different types of scaffolds

The use of PCK-scaffolds increased the diagnostic competences of the pre-service teachers, showing significant moderate to strong effects. As the tasks in the video-based learning environments can be assumed as complex and difficult for the pre-service teachers (Kramer et al., 2020), scaffolding is supposed to be a highly effective intervention (Belland et al., 2017; Quintana et al., 2004). During the intervention, the pre-service teachers learn through the scaffolds and construct their knowledge with the help of them. The context-specific scaffolds also put a focus on the relevant aspect of biology-specific instructional quality in the classroom situation (Wood et al., 1976), so that this kind of high guidance helps the learners with little prior knowledge to successfully complete the tasks (Chernikova et al., 2020). Even when pre-service teachers diagnosed classroom situations without the support of scaffolds, as it was the case when *DiKoBi^III^ Assess* was utilised in the post-test, pre-service teachers of the PCK-group were still able to successfully complete the tasks. Hence we can assume, that they did not only benefit from the scaffolds in the moment of their presentation but also in future tasks (Reiser, 2004). Even though the mere use of videos can already elicit professional knowledge (Seidel & Stürmer, 2014), context-specific scaffolds can help to focus the diagnosis on the essentials and thus increase the accuracy of the diagnosis.

pre-service teachers receiving both scaffolds, PCK-scaffolds and the DA-scaffolds (PCK + DA), also showed a significant improvement from pre- to post-test. Although they received the PCK-scaffold and thus are prompted with the relevant information necessary to perform the three tasks, the increase in the number of keywords used was not as high as for the PCK-group. As a consequence of including the DA-scaffolds as second type of scaffold in the learning environment, the time pre-service teachers needed to read the scaffolds and the working time increased. According to Moosbrugger and Kelava (2012), the motivation decreases as the learning environment progresses, and thus, the processing time increases. Furthermore, the cognitive load might be higher for this group due to the amount of text in two instead of one scaffold (Leahy & Sweller, 2016). In conclusion, pre-service teachers receiving context-specific scaffolds in the intervention, like the PCK-group and the PCK + DA-group, showed better diagnostic competences in the post-test than in the pre-test.

The diagnostic competences of the DA-group were not significantly higher in the post-test than in the pre-test. All pre-service teachers are in an early stage of their university education. Therefore, it can be assumed that their prior pedagogical content knowledge is rather low and that pre-service teachers may benefit from additional declarative information on the content to successfully complete the tasks. In contrast, a scaffold that provides action-related information about the diagnostic activities is intended to support diagnosing and is therefore not suitable for learners with little prior knowledge (Chernikova et al., 2020). The three tasks in *DiKoBi* can be interpreted as scaffolds themselves, so that adding a DA-scaffold such as the one shown in Table 3 does not bring up something new to help the pre-service teacher to complete the task and can be categorised as redundant. However, the learning environment *DiKoBi* contained tasks that prompted the diagnostic activities already to a certain amount, e.g. for the task *explain*: 'Give reasons why you think your observed aspects of teaching need improvement. Try to use subject-related biology-specific instructional theories to justify your observation'. The *redundancy effect* as described by Chandler and Sweller (1991) occurs when the same information is given twice but in different forms of presentation. The learner then must process the fact, that the information of the two sources is identical what increases the cognitive load and thus makes learning less effective than if just one of the sources would be presented (Sweller et al., 2019).

The three DA scaffolds appeared in all tasks of the five classroom situations, so that the pre-service teacher reads the same three scaffolds five times. This repetitive scaffolding could maybe lead to a decreased motivation having a negative effect on the pre-service teachers' learning.

To explain the results concerning the DA-scaffolds, it is also important to keep in mind that we operationalised the diagnostic competences by the number of keywords used and that the DA-scaffolds in contrary to the PCK-scaffolds do not contain any of the keywords. The keywords were so not directly scaffolded and only the hint of including biology-instructional theories can be considered as prompt for the keywords. As the pre-service teachers have little prior knowledge, the group receiving scaffolds for the diagnostic activities, did not have higher scores in the post-test than in the pre-test.

The correlation between the number of keywords used and the sum score was remarkably high. The keywords which we included in the scaffolds represented the biology-specific instructional quality aspects that were demanded for diagnosing the classroom situations. Thus, the high correlation indicates that a correct and applicable diagnosis can only be made using the keywords to describe and explain the problem. Besides, pre-service teachers automatically get two out of three point for mentioning the addressed concept in the task *explain* (Kramer et al., 2020) and these concepts often contain the keywords.

With regard to RQ1, it can be summarised that only the use of context-specific scaffolds, but not generic scaffolds, could improve the diagnostic competences of pre-service biology teachers in the video-based learning environment *DiKoBi*.

6.2. Differences between the different types of scaffolds

We could investigate significant differences between the four groups in the post-test. In the pre-test, there were no statistically significant differences between the four groups, whereas in the intervention differences became significant. The PCK group achieved the highest number of keywords, followed by the PCK + DA group. Both groups significantly differ from the control group and the PCK-group even showed significance towards the DA-group. These results were little surprising, seeing that the keywords are represented directly when pre-service teachers receive the PCK-scaffold. More importantly, the results of the post-test should be considered. Comparing the four groups, two groupings can be formed: (1) the PCK- and the PCK + DA-group, (2) the DA- and the control group. They differed in terms of the type of scaffolds pre-service teachers received. There was significant improvement for pre-service teachers receiving a content-related PCK-scaffold and there was no significant improvement when they did not receive that scaffold. This can be approved with results from Chernikova et al.'s (2020) meta-analysis constating that learners with little prior knowledge benefit the most of scaffolds providing additional information and that scaffolds supporting the process of diagnosing are not effective. The PCK-group was scoring the highest. The PCK + DA-group scored higher than the DA-group but lower than the PCK-group. This can be attributed to the length of the text pre-service teachers had to read and the associated higher cognitive load (Leahy & Sweller, 2016). The standard deviation in the PCK- and in the PCK + DA-group is remarkably high during the intervention. A possible explanation for this might also be the increased cognitive load as it could be overloading for some pre-service teachers to read the rather long PCK-scaffold in addition to the video and for others not so much. The DA-group did not use significantly more keywords than the control group neither in the intervention nor in the post-test. This leads to the assumption, that the DA-scaffold did not support pre-service teachers (with a low level of prior knowledge) to successfully complete the tasks neither when they are presented in the moment nor did the scaffolds support the learning process for future tasks. Three rather extreme outliers can be found in the PCK + DA-group during the intervention (see boxplots in Figure 4). All calculations were re-run without those outliers and these analyses showed very similar results and the effects did not vary dramatically. Thus, we decided to consider the outliers as part of the data set.

6.3. Limitations

In addition, there are some limitations in our study. As the learning environment *DiKoBi* consists of five videos showing the classroom situations and the three tasks *Describe*, *Explain* and *Alternative Strategy* for each situation, it takes a relatively long time to complete it. This can decrease motivation from classroom situation one to classroom situation five (Moosbrugger & Kelava, 2012). Going through the whole learning environment three times in total (pre-test, intervention and post-test), pre-service teachers may also experience motivational effects in form of test fatigue (Döring & Bortz, 2016). Nevertheless, we could investigate a higher score for all groups in the post-test, so that there is no drastic drop in motivation and the effect of learning is superior to possible motivational effects. Concerning the prior knowledge of the pre-service teachers, we did not directly assess the pre-service teachers' PCK, but assumed it based on the number of courses they took in biology instruction.

Despite the small sample size that can affect the validity of the results (Döring & Bortz, 2016), we achieved significant results, and thus, can derive generalisations.

6.4. Conclusions

In conclusion, pre-service teachers need a minimum level of PCK to be able to diagnose biology-specific aspects of instructional quality in the video-based learning environment

DiKoBi. If pre-service teachers have little prior knowledge, content-related PCK-scaffolds can support the learning process and effectively foster diagnostic competences. Generic scaffolds that only support the process of diagnosing (DA-scaffolds in this case) cannot foster diagnostic competences for pre-service teachers with little prior knowledge. These pre-service teachers do not show better results than the control group. To successfully go through the process of diagnosing in a specific context (e.g. biology instruction), learners thus first need a certain amount of prior knowledge in the specific domain of the context (e.g. PCK) before diagnostic competences can be fostered. Applying the results on the two components of diagnostic competences (Heitzmann et al., 2019), we can summarise, that the knowledge in the specific domain, which here is PCK, seems to be crucial for a successful diagnosis in teaching contexts.

Following on from this, it would be interesting for further studies, to improve the integration of scaffolds in the learning environment *DiKoBi* by adapting PCK-scaffolds to the prior knowledge of the pre-service teachers and to effectively support pre-service biology teachers in their learning process and help them construct diagnostic competences.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The current study was funded by grants of the German Research Foundation (DFG) (Grant: NE 1196/8-1, FOR 2385).

ORCID

Marie Irmer http://orcid.org/0000-0002-2509-0706
Dagmar Traub http://orcid.org/0000-0001-6618-1084
Maria Kramer http://orcid.org/0000-0003-4243-2867
Christian Förtsch http://orcid.org/0000-0002-3625-6791
Birgit J. Neuhaus http://orcid.org/0000-0001-6031-9660

References

Arnold, J., Kremer, K., & Mayer, J. (2014). Schüler als Forscher – Experimentieren kompetenzorientiert unterrichten und beurteilen [Pupils as researchers – teaching and assessing experimentation in a competence-oriented way]. *Mathematisch und Naturwissenschaftlicher Unterricht*, *67* (2), 83–91.

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

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*, *25*(1), 307–316. https://doi.org/10.1007/s40573-019-00103-9

Belland, B. R. (2017). Computer-based scaffolding strategy. In B. R. Belland (Ed.), *Instructional scaffolding in STEM education* (pp. 107–126). Springer International Publishing. https://doi.org/10.1007/978-3-319-02565-0_5

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

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research*, *87*(2), 309–344. https://doi.org/10.3102/0034654316670999

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

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Reseacher*, *33*(8), 90–114. https://doi.org/10.3102/0013189X033008003

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

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*(1), 157–196. https://doi.org/10.1007/s10648-019-09492-2

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, *90*(4), 499–541. https://doi.org/10.3102/0034654320933544

Cho, K-L, & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development*, *50*(3), 5–22. https://doi.org/10.1007/BF02505022

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship. *Thinking: The Journal of Philosophy for Children*, *8*(1), 2–10. https://doi.org/10.5840/thinking19888129

Dorfner, T., Förtsch, C., Boone, W., & Neuhaus, B. J. (2019). Instructional quality features in videotaped biology lessons: Content-independent description of characteristics. *Research in Science Education*, *49*(5), 1457–1491. https://doi.org/10.1007/s11165-017-9663-x

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 a concept-oriented biology lesson? A planning model for biology lessons - The shell model.]. *MNU Journal*, *72*(4), 300–306.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften [Research methods and evaluation in the social and human sciences]*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41089-5

Field, A. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)* (3rd ed.). Sage Publication. http://www.uk.sagepub.com/field3e/main.htm

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B. J., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J-W, Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *2*(3), 28–45. https://doi.org/10.14786/flr.v2i2.96

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., Kotzebue, L. v., & 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

Gartmeier, M., Bauer, J., Fischer, M. R., Hoppe-Seyler, T., Karsten, G., Kiessling, C., Möller, G. E., Wiesbeck, A., & Prenzel, M. (2015). Fostering professional communication skills of future physicians and teachers: Effects of e-learning with video cases and role-play. *Instructional Science*, *43*(4), 443–462. https://doi.org/10.1007/s11251-014-9341-6

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

Glogger-Frey, I., Herppich, S., & Seidel, T. (2018). Linking teachers' professional knowledge and teachers' actions: Judgment processes, judgments and training. *Teaching and Teacher Education*, *76*(1), 176–180. https://doi.org/10.1016/j.tate.2018.08.005

Grossmann, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record: The Voice of Scholarship in Education*, *111*(9), 2055–2100. https://doi.org/10.1177/016146810911100905

Hannafin, M., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods, and models. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (pp. 115–140). Lawrence Erlbaum Associates, Inc.

Heitzmann, N. (2014). *Fostering diagnostic competence in different domains* [München. Ludwig-Maximilians-Universität. 2014, Universitätsbibliothek der Ludwig-Maximilians-Universität, München]. Deutsche Nationalbibliothek.

Heitzmann, N., Seidel, T., Hetmanek, A., Wecker, C., Fischer, M. R., Ufer, S., Schmidmaier, R., Neuhaus, B., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., Fischer, F., & Opitz, A. (2019). Facilitating diagnostic competences in simulations in higher education: A framework and a research agenda. *Frontline Learning Research*, *7*(4), 1–24. https://doi.org/10.14786/flr.v7i4.384

Jahn, G., Stürmer, K., Seidel, T., & Prenzel, M. (2014). Professionelle Unterrichtswahrnehmung von Lehramtsstudierenden [Professional teaching perception of student teachers]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *46*(4), 171–180. https://doi.org/10.1026/0049-8637/a000114

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*, *68*, Article 100973. 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 Journal*, *7*(1). https://doi.org/10.1080/2331186X.2020.1823155

Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science*, *44*(1), 107–123. https://doi.org/10.1007/s11251-015-9362-9

Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung [Competence model of scientific acquirement of knowledge.]. In U. Harms & A. Sandmann (Eds.), *Lehr- und Lernforschung in der Biologiedidaktik* (pp. 63–79). Studienverlag.

Meisert, A. (2012). Wie kann Biologieunterricht geplant werden? [How to plan biology lessons?]. In U. Spörhase-Eichmann & K. Köhler (Eds.), *Biologie-Didaktik: Praxishandbuch für die Sekundarstufe l und ll* (5th ed.) (pp. 241–274). Cornelsen.

Moosbrugger, H., & Kelava, A. (2012). *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion. [Springer textbook. Test theory and questionnaire construction]*. Springer. http://d-nb.info/101263051x/04

Nodari, C., & Steinmann, C. (2008). Fachdingsda: Fächerorientierter Grundwortschatz für das 5.-9. Schuljahr (1. Aufl.). [Fachdingsda: Subject-oriented basic vocabulary for the 5th-9th school year (1st ed.)] *Lehrmittel der Interkantonalen Lehrmittelzentrale.* Lehrmittelverlag des Kantons Aargau.

Nowak, K. H., Nehring, A., Tiemann, R., & Upmeier zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education*, *47*(3), 182–188. https://doi.org/10.1080/00219266.2013.822747

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). Teachers' diagnostic competences and their practical relevance. *Journal of Educationals Research Online*, *7*(2), 5–10. https://doi.org/10.25656/01:11487

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275–300. https://doi.org/10.1080/15391523.2020.1719943

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design Framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(3), 337–386. https://doi.org/10.1207/s15327809jls1303_4

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304. https://doi.org/10.1207/s15327809jls1303_2

Riebling, L. (2013). Sprachbildung im naturwissenschaftlichen Unterricht: Eine Studie im Kontext migrationsbedingter sprachlicher Heterogenität (1. Aufl.) [Language education in science teaching: A study in the context of migration-related linguistic heterogeneity (1st ed.)]. In *Interkulturelle Bildungsforschung* (p. 20). Waxmann Verlag GmbH.

Roth, K. J., Bintz, J., Wickler, N. I. Z., Hvidsten, C., Taylor, J., Beardsley, P. M., Caine, A., & Wilson, C. D. (2017). Design principles for effective video-based professional development. *International Journal of STEM Education*, *4*(1), 31. https://doi.org/10.1186/s40594-017-0091-2

Santagata, R., & Guarino, J. (2011). Using video to teach future teachers to learn from teaching. *ZDM*, *43*(1), 133–145. https://doi.org/10.1007/s11858-010-0292-3

Schmelzing, S., van Driel, J. H., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-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. https://doi.org/10.1007/s10763-012-9384-6

Schmiemann, P., Linsner, M., Wenning, S., Neuhaus, B. & Sandmann, A. (2011). Kontextorientiertes Lernen in Biologie – Aufgaben und Arbeitsmaterialien [Context-oriented learning in biology – tasks and working materials.] In P. Schmiemann & A. Sandmann (Hrsg.), Aufgaben im Kontext: Biologie (p. 4–12). Seelze: Friedrich Verlag.

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

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Reseacher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science*, *35*(1), 41–72. https://doi.org/10.1007/s11251-006-9002-5

Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der simulierte Klassenraum [The simulated classroom]. *Zeitschrift Für Pädagogische Psychologie*, *22*(3-4), 293–309. https://doi.org/10.1024/1010-0652.22.34.261

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Upmeier zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht [Model competence in biology teaching]. *Zeitschrift Für Didaktik Der Naturwissenschaften*, *16*, 41–57.

Voss, T., & Kunter, M. (2013). Teachers' general pedagogical/psychological knowledge. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 207–227). Springer. https://doi.org/10.1007/978-1-4614-5149-5_10

Vygotskij, L. S. (1979). Interaction between learning and development. In M. Cole & L. S. Vygotskij (Eds.), *Mind in society: The development of higher psychological processes* (2nd ed.) (pp. 79–91). Harvard Univ. Press.

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*(1), 25–47. https://doi.org/10.1007/s10763-012-9390-8

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x