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

Maria Kramer1*, Christian Förtsch1, Julia Stürmer1, Sonja Förtsch1, Tina Seidel2 and Birgit J. Neuhaus1

Abstract: A key challenge for teachers is the evaluation of teaching and learning situations in the classroom. Video-based instruments are considered as effective tools for measuring teachers’ skills for evaluating classroom situations. This article reports the development and validation of a video-based assessment tool for measuring biology teachers’ professional vision. Analyses of interviews using think-aloud protocols with in-service biology teachers showed that they perceived the staged videos in the video-based assessment tool as authentic. Furthermore, they identified biology-specific dimensions of instructional quality in the videos, and used aspects of professional vision while reasoning about the identified situations. Thus, results indicate the suitability of the video-based assessment tool for capturing professional vision with the potential to be used in further interventions within teacher education programs.

Subjects: Teaching & Learning; Teachers & Teacher Education; Classroom Practice; Science Education

Keywords: Science education; teaching effectiveness; instructional quality; professional vision; situation-specific skills; authenticity; video-based assessment

ABOUT THE AUTHORS

Maria Kramer is a doctoral candidate at the Biology Education, LMU Munich, Germany. Her main research focuses on biology teacher education and professional competence of biology teachers. As a member of the research team COSIMA, she investigates the facilitation of diagnostic competences in simulation-based learning environments in higher education.

Christian Förtsch is a post-doctoral researcher at the Biology Education, LMU Munich, Germany. He is interested in research on instructional quality of biology lessons as well as on biology teachers’ professional competence, including professional knowledge, professional vision, diagnostic competences, and digital skills.

Birgit J. Neuhaus is Professor in Biology Education at LMU Munich, Germany. Her research interests are mainly connected to biology teachers’ professional competence and instructional quality of biology lessons.

PUBLIC INTEREST STATEMENT

A crucial ability of a science teacher is that he or she is able to identify subject-specific problems in the classroom quickly and to make appropriate decisions that result in a good teaching practice, and thus, to improve students’ achievement. The ability to notice such problems and interpret them appropriately is called professional vision. Professional vision is considered an important key characteristic of teachers, and thus, should be promoted already in university education. To assess the effectiveness of supportive programs, subject-specific measuring instruments are needed. The use of video-based tools is considered to be particularly promising, since videos can be used to measure professional vision in a situated context of real-world demands. This article reports the development and validation of a video-based learning environment that in future might fill the gap between theory and practice within university science teacher education.
1. Introduction

Effective teaching can be seen as knowing about instructional quality and its implementation in the classroom. To evaluate effective teaching, systematic observation of teaching components is considered as crucial part of teachers’ professional competence. Systematic observation of teaching components includes observation and assessment of teachers’ behavior, which results from knowledge teachers draw on in the moment of action (Alonzo et al., 2019; Biggs & Tang, 2007; Sherin, 2002). The concept of professional vision followed this approach of investigation, for example, by using videos. Professional vision can be seen as mediating situation-specific skill between cognitive dispositions and classroom performance. Teachers’ professional knowledge, as part of cognitive dispositions, influences their professional vision that in turn affects their performance in the classroom and, therefore, instructional quality of their teaching (Blömeke et al., 2015; Goodwin, 1994; Meschede et al., 2017).

It is assumed that teachers’ professional competence develops bi-directionally, because their knowledge and their teaching performance interact with each other while planning, teaching, and reflecting upon their instruction (Blömeke & Kaiser, 2017; Santagata & Yeh, 2016). Because of the bi-directionality of the competence development, biology teachers’ application of knowledge during their assessment of classroom situations brings an opportunity to support them in developing their knowledge. This approach is also used within the Refined Consensus Model, a framework, which illustrates the circular shaping process of teachers’ subject-specific knowledge during planning, teaching, and reflecting (Alonzo et al., 2019). Since the assessment of classroom situations depends on the subject-specific professional knowledge of teachers, their pedagogical content knowledge as well as their subject-specific professional vision must be taken into account (Steffensky et al., 2015).

Consequently, science teachers should have various opportunities to apply their subject-specific knowledge to teaching, gain experiences, reflect on their teaching, and develop their routines of teaching (Kane et al., 2013; Santagata & Yeh, 2016). In order to provide classroom situations, in which subject-specific knowledge can be applied to assess science teaching, we developed the video-based assessment tool DiKoBi Assess (German acronym for diagnostic competences of biology teachers in biology classrooms). Here, a biology teacher’s instruction on a specific biological content, and thus, subject-specific instructional quality in a real-life teaching situation can be assessed by other biology teachers who act as observers. The video-based assessment tool offers opportunities for biology teachers to put their knowledge into practice. This article reports the validation of the content of the videos used in DiKoBi Assess with regard to their authenticity and the recognizability of the integrated challenges (part A) as well as the validation of the assigned tasks with regard to their suitability for capturing aspects of professional vision (part B).

2. Theoretical background

2.1. Effective teaching

The behavior of a teacher in the classroom is considered to be decisive for teaching effectiveness. Research on teaching effectiveness has described generic as well as subject-specific features of instructional quality, which influence student achievement (e.g., Brunner et al., 2006; Hattie, 2009; Kyriakides et al., 2013; Seidel & Shavelson, 2007; Wüsten, 2010). In German-speaking countries, generic features have been summarized into three basic dimensions of instructional quality. Similar approaches can also be found in English-speaking countries, such as the CLASS-S framework (Pianta et al., 2012). The three basic dimensions are classroom management, supportive climate, and cognitive activation (Baumert et al., 2010; Praetorius et al., 2018). Classroom management describes how a teacher structures, organizes, and monitors teaching and learning situations, and how he or she deals with student disruptions and establishes rules and routines to increase time on task and
students’ interest in the subject (Kunter et al., 2007; Praetorius et al., 2018; Seidel & Shavelson, 2007). The basic dimension supportive climate contains features such as a positive relation between teachers and students as well as constructive feedback that benefits the learning atmosphere in the classroom (Praetorius et al., 2018). Cognitive activation refers to the facilitation of a deeper understanding of the content and the consideration of prior knowledge and ideas of students. Since cognitive activation cannot be observed directly, it is often indirectly described by the teachers’ instruction. Important subdimensions of cognitively activating instruction are the level of students’ cognitive activities, and conceptual instruction. In order to analyze classroom situations for evidence of these subdimensions of cognitive activation, elements such as exploring and activating prior knowledge, using challenging tasks and questions, or supporting knowledge linking have been used (Förtsch et al., 2017; Lipowsky et al., 2009; Praetorius et al., 2018).

With regard to interrelations between the three basic dimensions, Dorfner, Förtsch, Neuhaus et al. (2018) underlined the establishment of good classroom management and supportive climate as prerequisites for implementing cognitively activating strategies. Furthermore, cognitive activation is the only basic dimension, in which subject-specificity is considered (Dorfner et al., 2017; Schlesinger & Jentsch, 2016). Dorfner, Förtsch, Boone et al. (2019) showed the suitability of the basic dimensions for content-independent descriptions of instructional quality, but they also emphasized the need of features that are closer to the specificity of teaching a particular subject, especially science, of which they discussed several subject-specific features of instructional quality. Table 1 shows specific features of instructional quality that can extend the basic dimensions for science education. These features, exemplified for biology instruction, were found to be empirically effective and can be implemented in the course of a lesson (Dorfner, Förtsch, Spangler et al., 2019).

As part of teachers’ professional competence, teachers need to know about the subject-specific instructional quality features and transfer them to content-specific situations of teaching and learning. For effective biology teaching, teachers have to pay attention to biology-specific features of instruction and implement them in their classroom practice (Biggs & Tang, 2007; Steffensky & Neuhaus, 2018). A crucial part of biology teachers’ expertise is their ability to identify effective biology teaching as well as to reason about the identified situations in an evidence-based and concept-oriented manner (Kane et al., 2013; Santagata & Yeh, 2016). Promoting professional competence of teachers is therefore an important element in their development of effective teaching practice.

### 2.2. Teachers’ professional competence

Blömeke et al. (2015) modelled professional competence on a continuum between dispositions and performance, mediated by situation-specific skills. The model can be applied to varying situations such as evaluating effective teaching. Cognitive dispositions such as professional knowledge and affective-motivational dispositions such as beliefs influence the application of situation-specific skills within instruction, and thus, foster student outcomes (Baumert & Kunter, 2013; Blömeke & Kaiser, 2017). Teachers need perceptual skills to identify characteristics of classroom situations; they have to analyze and interpret them, and finally make a decision on using appropriate teaching strategies. In teacher education, the ability to notice and reason about relevant classroom situations is discussed in terms of the concept of professional vision (Goodwin, 1994; Van Es & Sherin, 2002). Professional vision requires conceptual knowledge of effective teaching as well as the ability to transfer this knowledge in form of interpretations to the situation observed (Borko, 2004; Van Es & Sherin, 2008). Thus, Stürmer and Seidel (2015) considered professional vision as a situation-specific skill that combines knowledge and practice.

#### 2.2.1. Role of professional knowledge

Existing models have already described the structure of teachers’ professional knowledge. Based on the work of Shulman (1986), professional knowledge of teachers can be divided into general pedagogical-psychological knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK) (cf. Baumert & Kunter, 2013; Blömeke et al., 2014). PK includes interdisciplinary
Table 1. Subject-specific dimensions and related features that define instructional quality in science education, exemplified for biology lessons

| Subject-specific dimension | Description of subject-specific features and empirical evidence                                                                                                                                                                                                 | Biology-specific example on the topic skin                                                                                                                                                                                                 |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Level of students' cognitive activities | It describes cognitively demanding instruction that fosters students' deep understanding of the content. Cognitive activities can be assessed with regard to a single task or the whole teaching process and its structure. Indicators are the 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 (Förtsch et al., 2016, 2017; Lipowsky et al., 2009; Nawani et al., 2017; Praetorius et al., 2018). | Lesson starts with a cognitively stimulating component: comparison of a knight’s armor with human skin.                                                                                                                                                                                                 |
| Conceptual understanding | Instruction should include high knowledge linking to foster students’ knowledge structure. By teaching interrelated facts and concepts instead of isolated facts, students are cognitively activated and thus the development of their knowledge structure can be fostered. Knowledge linking also plays a central role, when students’ prior knowledge and ideas are activated and linked with already known concepts, as well as when knowledge is applied and transferred to multiple contexts (Förtsch et al., 2020; Lipowsky et al., 2009; Steffensky & Neuhaus, 2018; Wadouh et al., 2014). | At the end of the lesson, the teacher refers back to the initial comparison of a knight’s armor with human skin. Students have to apply their acquired knowledge on this context.                                                                                                                                 |
| Creation of situational interest | Situational interest can be understood as a psychological state that involves conditions like focused attention and affective involvement. Subject-specific characteristics like referring to everyday examples or experiments can enhance intrinsic motivation and situational interest (Dorner, Förtsch, Neuhaus et al., 2018; Schiefefe, 2009; Todt, 1977). | The comparison of a knight’s armor with human skin functions as a catch that creates focused attention and affective involvement.                                                                                                                                                                                                                       |
| Dealing with students’ ideas and errors | An appropriate way of dealing with students’ errors means to detect technical errors and use these errors as an opportunity to their learning. Examples of a formative handling are: listing errors as assumptions on the board and discuss them later in class, encourage students to explain their ideas and solution methods, or simply correct the errors right away. The aim is to change students’ misconceptions (Herppich et al., 2013; Rach et al., 2013; Spychiger, 2008). | The teacher uses a misconception of a student (e.g., the skin notices when you are touched) as assumption that is noted at the board. After working on the topic, students can discuss the misconception and correct it.                                                                                                                                                  |

(Continued)
| Subject-specific dimension | Description of subject-specific features and empirical evidence | Biology-specific example on the topic skin |
|----------------------------|----------------------------------------------------------|-----------------------------------------|
| Use of technical language  | Using minimally necessary topic-relevant terms in biology instruction with additional use of biological core ideas can positively affect students’ learning of conceptual knowledge. The overall number of technical terms should be reduced. Explaining and understanding new concepts in everyday terms prior to using scientific language supports students’ conceptual and linguistic understanding (Brown & Ryoo, 2008; Dorfner, Förtsch, Neuhaus et al., 2019). | The teacher explains the structure and function of sweat glands without mentioning all technical terms for the different types of glands that exist. |
| Use of experiments         | To solve problems, scientific inquiry methods, such as experiments, can be used. Following the concept of scientific inquiry, process variables can be described. These variables have to be implemented during experimentation to foster scientific reasoning. Hence, formulating scientific research questions and hypotheses, planning experiments, analyzing data, and drawing conclusions are seen as necessary, cognitively activating process variables when using experiments in instruction (Dorfner, Förtsch, Germ et al., 2018; Mayer, 2007; Simon, 1992; Tesch & Duit, 2004). | Steps of scientific inquiry are embedded in the biology lesson. For example, to understand the specific sensibility of different skin areas, first, the students hypothesize the initial phenomenon, second, they plan and conduct an appropriate test, and finally they interpret their results. |
| Use of models              | An elaborate use of models (e.g., by using models as tools for scientific reasoning) can foster students’ learning. It is assumed that using models for explaining learning content rather than just for illustration fosters students’ achievement and scientific reasoning skills. When using models on an elaborate level, students need to think about the content more deeply, which is considered more cognitively activating. A critical reflection of the model as representation of something is also part of the modelling process (Förtsch et al., 2017; Oh & Oh, 2011). | Three different models showing the profile of different areas of the skin (fingertip, upper arm, back) are used as a method of scientific inquiry. Additionally, the models are critically reflected and compared with the original (that is our skin). |

Note. References to associated concepts as well as research findings are added after each description of subject-specific features.

knowledge about several teaching methods, learning strategies, and classroom management. CK is described as knowledge about subject-specific content and its conceptual understanding; and PCK is knowledge that teachers need to make this content accessible to their students (Fischer et al., 2012; Shulman, 1986; Voss & Kunter, 2013). Even though different conceptualizations of PCK exist, they all take into account knowledge about students’ misconceptions, knowledge about subject-specific structures of biology instruction, and corresponding teaching strategies (Depaepe
et al., 2013; Schmelzing et al., 2013). When enacting instruction in the classroom, teachers need to use knowledge from all facets. Thus, CK, PCK, and PK are shown to be determinants of instructional quality, which influence students’ learning outcomes and motivational development (Baumert et al., 2010; Grossman & McDonald, 2008; Förls & et al., 2016; Voss & Kunter, 2013).

In terms of biology teachers’ professional competence, the subject-specific knowledge dimension PCK is considered to be a determining factor for effective biology teaching (Förls & et al., 2016). To measure science teachers’ PCK, researchers used methods such as classroom observations or semi-structured interviews, they collected and analyzed lesson plans, students’ work samples, teachers’ written reflections, as well as researcher’s field notes. Others developed tools included Content Representation (CoRe) and Pedagogical and Professional Experiences Repertoires (PaP_eRs) for documenting and portraying PCK; and some studies also used a combination of these methods in order to form a genuine representation of the complex construct of PCK (cf. Barendsen & Henze, 2019; Lee & Luft, 2008; Lehan & Bertram, 2016; Lougahan et al., 2004; Park & Oliver, 2008). Over the past years, researchers discussed different aspects of PCK that may form this complex construct. For example, Alonzo and Kim (2016) emphasized the differentiation between declarative PCK (knowing that) and more dynamic forms of PCK “that cover teachers’ activities during a lesson, for example, if a teacher is able to react appropriately to students’ questions and mistakes” (Schmelzing et al., as cited in Alonzo & Kim, 2016, p. 1260).

The acquisition of PCK and its transfer to effective teaching performance is described as a circular shaping process, in which knowledge and practical experience influence each other (Blömeke & Kaiser, 2017; Santagata & Yeh, 2016). The Refined Consensus Model (RCM) (Carlson & Daehler, 2019) is an attempt to represent this process in a modeling way. The RCM distinguishes between three different forms of PCK that develop and interact during teachers’ professionalization: collective PCK (cPCK), personal PCK (pPCK), and enacted PCK (ePCK). The general canonical knowledge someone needs to know to teach biology represents cPCK. This form is held by multiple educators of biology education, and collected and reported in the literature on biology education (Carlson & Daehler, 2019). In contrast, pPCK and ePCK are more individual forms. In this vein, pPCK describes the personal knowledge resulting from the teaching and learning experiences of an individual biology teacher and his/her reflection on such experiences. pPCK interacts largely with ePCK, which is considered to be the core of teachers’ professional competence that only occurs in the moment of acting in biology classrooms, and therefore, has a rather tacit nature (Alonzo et al., 2019; Sorge et al., 2019). Through reflection in and on action, the generated ePCK can be transformed into pPCK afterwards (Alonzo et al., 2019; Schön, 1987). By contributing to pPCK and sharing it with a group of biology teachers, aspects of pPCK can also be transferred into cPCK and thus, can complement the canonical knowledge of the domain.

Considering ePCK as a dynamic form including skills and knowledge to plan (ePCKplan), teach (ePCK teach), and reflect (ePCK reflect) on instruction and student outcomes within the moment of acting, the proximity to Blömeke and Kaiser (2017) conception of situation-specific skills that teachers use for planning, teaching, and reflecting might come to mind. The crucial difference between these two conceptions of teacher knowledge is the explicit consideration of tacit knowledge within the RCM (Alonzo et al., 2019), whereas Blömeke and Kaiser (2017) did not mention tacit knowledge in their definition of situation-specific skills. The emphasis of the tacit nature of ePCK also raises the question of how to measure this form of PCK valid and reliable. Due to the definition of ePCK as tacit knowledge and the contextualized nature thereof, researchers demand other approaches of inferring or approximating ePCK. One way to approximate ePCK might be seen in the conception of situation-specific skills. A situation-specific skill that researchers used for investigating reflection on instruction is professional vision.
2.2.2. Professional vision as central part of teachers’ competence

The transfer of professional knowledge to high-quality instruction in real-life performance depends on the execution of situation-specific skills to perceive events relevant to learning, to interpret them, and decide on pursuing strategies. Transferring the RCM to the competence as a continuum model, the execution of situation-specific skills can be considered as activation and generation of ePCK during teachers’ planning of instruction, teaching during instruction, or reflecting on instruction (Carlson & Daehler, 2019).

Based on the situation-specific skill professional vision, which is described as “socially organized ways of seeing and understanding events that are answerable to the distinctive interests of a particular social group” (Goodwin, 1994, p. 606), noticing and reasoning are described as two components for professionally observing and interpreting classroom situations (Borko, 2004; Seidel & Stürmer, 2014; Van Es & Sherin, 2002). Whereas these components are characterized by explicit elements, which are measurable, there are also indications that they can contain tacit elements, which have an impact on teachers’ noticing and reasoning (Sherin et al., 2008). Therefore, if we understand observing and interpreting classroom situations as a way of reflecting on instruction, the situation-specific skill professional vision might be considered as approximation to ePCK_{reflect}, but not as exact representation.

Seidel and Stürmer (2014) described noticing as paying attention to events that are relevant for students learning, and reasoning as knowledge-based interpretation of classroom situations including the three aspects: description, explanation, and prediction. Description means referring to “relevant aspects of noticed teaching situations and learning components without making further judgements” (Seidel & Stürmer, 2014, p. 745); explanation means to use someone’s knowledge to reason about the noticed aspects by linking them to concepts and theories; and prediction means to derive consequences from “observed events in terms of student learning” (Seidel & Stürmer, 2014, p. 746). In other words, explaining and predicting helps teachers to come to pedagogical decisions. These decisions can also occur as suggestions of alternative responses to student answers or alternative teaching strategies that can make teaching more effective (Santagata & Yeh, 2016; Van Es & Sherin, 2002). As the aspects of professional vision are perceived to develop over time, teacher education has to support teachers in gaining them as part of their professional development. An increasingly used method for supporting teachers’ professional vision is video viewing (Gaudin & Chaliètes, 2015).

2.3. Video use in teacher education

Professional competence can be measured with different instruments (e.g., Baumert et al., 2010; Blömeke et al., 2014). When measuring professional vision as part of teachers’ professional competence, a video-based approach is promising. Video-based instruments are considered as appropriate to foster the transfer of knowledge about effective teaching and learning principles to practice, and thus, they are considered as a potentially effective tool for professional development (Roth et al., 2017; Seidel & Stürmer, 2014; Shavelson, 2012; Stürmer et al., 2013; Treisch, 2018).

In teacher education, the number of recently designed video-based instruments for assessing professional vision or similar constructs with general or domain-specific perspectives is growing (e.g., Gold & Holodynski, 2017; Kersting et al., 2010; König & Kramer, 2016; Meschede et al., 2017; Michalsky, 2014; Seidel et al., 2010; Wiens et al., 2020). These video-based tools have in common that skills such as professional vision are captured using task prompts. By completing the tasks given in the instrument, teachers’ ability to apply knowledge about teaching and learning within videotaped classroom situations can be measured (Seidel & Stürmer, 2014).

Besides video-based assessment tools, real-life classroom videos are used in video clubs to observe and interpret videotaped classroom situations as part of pre-service teachers’ higher education. Videos work as item prompts to evoke professional knowledge that can be applied for reasoning about instructional quality, and thus, can improve both professional knowledge and
reasoning skills (Behling et al., 2019; Roth et al., 2011; Santagata & Guarino, 2011; Star & Strickland, 2008; Van Es & Sherin, 2008). Referring to the RCM, watching classroom situations and reflecting upon them might raise teachers’ ePCK and pPCK, and thus gives opportunity to increase teachers’ situation-specific skills as well (Alonzo et al., 2019; Meschede et al., 2017). Additionally, researchers also emphasized the importance of video in raising students’ awareness of practices and actions that facilitates transformation (Roth et al., 2011; Siry & Martin, 2014).

Besides the use of videos that recorded real-life classroom situations, scripted (staged) classroom scenarios become more and more relevant, as they can be tailored to specific purposes and observation foci, and thus, can provide very condensed situations. Research underlined that staged videos are considered as an appropriate method to measure teachers’ expertise to analyse complex classroom situations. Similar to original real-life performance, staged scenarios can activate teachers’ professional knowledge and teaching strategies by the school context shown in the videos (Gaudin & Chaliès, 2015; Hoth et al., 2018; Kersting, 2008). In order to enable teachers to transfer acquired knowledge and skills to real-world situations, the shown classroom practice must be perceived as authentic. “Authentic” means that the situation shown in the video relates to real-life teaching contexts, requires skills of the discipline used similarly by more experienced teachers, and allows immersion in the situation (Barab et al., 2000; Belland et al., 2013; Messick, 1994b; Seidel et al., 2010). Therefore, a sufficient level of presence is also of considerable importance (Schubert et al., 2001). Overall, previous research indicates that staged videos of classroom situations are generally experienced as authentic (Piwowar et al., 2018).

So far, different components for achieving authenticity have been suggested. Components are: selecting key topic areas from national curricula to develop and review video situations that elicit rich discussion, ensuring an explicit focus of each video, and including mistakes to provide a true experience of exemplary teaching situations (Bliss & Reynolds, 2004; Kersting et al., 2010). To strengthen ecological validity, experts of the discipline (e.g., more experienced teachers) should be included in the development and review process of the practice representation and the instructional contexts, since experts are expected to know the cognitive processes and skills required for effective science teaching (Codreanu et al., 2020; Piwowar et al., 2018). Additionally, researchers emphasized that both the teacher and the students should be visible in the videos to enhance authenticity, and that the video situations are suggested to be relatively self-contained and short (Bliss & Reynolds, 2004; Kersting et al., 2010). Furthermore, authenticity can be increased by providing additional material that would be available in real-life situations as well. On the other hand, additional subject-specific information and material, subtitles or hints for task solution may cause distraction. Therefore, researchers emphasized to carefully select type and number of additional material but also the camera perspective to avoid heavy extraneous cognitive load that might decrease authenticity (Alonzo & Kim, 2016; Codreanu et al., 2020; Gold & Holodynski, 2017; Müller, 2016).

3. Research questions and objectives

Current research regarding evaluation of effective science teaching showed that subject-specific knowledge and professional vision are crucial aspects for instructional quality and teaching effectiveness (e.g., Kyriakides et al., 2013; Seidel & Shavelson, 2007; Steffensky et al., 2015). There is evidence that more practical experience that activates teachers’ ePCK positively affects the development of teachers’ professional competences, and thus, the development of their cognitive dispositions and cognitive skills as well (Stahnke et al., 2016). However, for better understanding of the interaction between ePCK generated by practical experience, and the other PCK-forms, research needs to define ways of measuring knowledge facets and forms validly. Practical approaches addressing subject-specificity of teachers’ professional competence might be useful, but need to be tested thoroughly (Van Es & Sherin, 2017; Gaudin & Chaliès, 2015).

Considering the subject-specificity of teaching effectiveness, researchers need valid instruments to systematically measure the structure and the development of professional vision with respect to
biology-specific, situated contexts of real-world demands (Blömeke et al., 2015). As far as we know, subject- and content-specific instruments for investigating biology instruction are still missing, but they are important to provide a necessary subject-specific lens for examining the development of biology teachers’ professional vision.

The aim of the study reported in this article is the development of a valid instrument to measure professional vision, and thus, approximate teachers’ ePCK. We assume that the development of skills for assessing classroom situations is highly subject-specific and necessary in teacher education as it prepares biology teachers to teach their subject effectively. To provide a tool, which represents real-world teacher practice for biology instruction, we developed the video-based assessment tool DiKoBi Assess. In our further studies, we want to use the biology-specific instrument to stimulate teachers’ reflection on teaching performances as part of their ePCK_reflect (Carlson & Daehler, 2019). In the first step, we tested the validity of the instrument with in-service teachers. Following a construct-centered approach (Messick, 1994b), our validation process was split into two parts, in which Part A referred to the validation of the content of the videos we used in DiKoBi Assess, and Part B to the validation of the tasks we assigned to trigger aspects of teachers’ professional vision. The following research questions were guidelines for our study.

Part A: Validation of the content of the videos

(RQ1) Do teachers assess the simulated classroom situations in the video as authentic?

(RQ2) Do teachers identify challenging classroom situations, which were derived from empirical studies and simulated in the videos?

Part B: Validation of assigned tasks

(RQ3) Do the given tasks within DiKoBi Assess measure the intended aspects of professional vision?

4. Methods
For measuring professional vision of biology teachers, we developed the video-based assessment tool DiKoBi Assess, and conducted interviews using think-aloud protocols with in-service teachers.

4.1. Development of the video-based assessment tool DiKoBi Assess
To develop a video-based assessment tool to investigate pre-service teachers’ professional vision, we recorded a set of staged video cases depicting biology-specific classroom situations, which were then embedded in an online-survey platform (Questback GmbH, 2018).

We scripted three consecutive lessons of the same class on the topic skin. Each lesson deals with a certain subtopic within the main topic skin and consists of six videos showing six different classroom situations (see Figure 1). In case participants use the assessment tool DiKoBi Assess more often, they do not have to watch the videos of the same lesson twice but can work on the classroom situations of one of the other lessons. Instruction of the scripted lessons followed standards from the Bavarian curriculum on the topic senses and sensory organs (State Institute of School Quality and Educational Research Munich, 2018). This topic can be taught at a similar level in both primary and early stages of secondary schools.

For each lesson, challenging aspects are shown in six classroom situations, each of which focuses on a different biology-specific dimension of instructional quality (in accordance with the dimensions listed in Table 1). The order of the situations follows the course of a lesson. The classroom situations include empirically proven, subject-specific dimensions of instructional quality as described in the theoretical section (see Section 2.1). In each classroom situation, relevant features of the focused dimension are missing. We consider these missing features as challenges that participants have to
identify when watching the videos (see Table 2). In order to assess the aspects of professional vision by using DiKoBi Assess, the tasks Describe, Explain, and Alternative Strategy were designed to optimally reveal the respective constructs (Messick, 1994a). Additional information—about sequence scheduling, the lesson plan, and the observation situation—was embedded before working on the video cases to minimize cognitive load during task solution (cf. Müller, 2016).

Before videotaping the staged video cases, scripts were prepared. Initially, we conducted a literature research about the topic skin. Additionally, the basic structure of the lessons in our scripts is based on lessons from teaching concepts of Förtsch et al. (2018) and Heidenfelder (2016) as well as on a lesson planning model for biology instruction (Dorffner, Förtsch, Spangler et al., 2019). We videotaped the staged video cases in school afternoon workshops at a secondary school in Bavaria. The students participated voluntarily in the workshop. Before starting the workshop, we obtained students’ and their parents’ written consent of participation in the workshop. The 90 minutes lasting workshop took place weekly. For videotaping, we used a special science classroom. This classroom was equipped with cameras for videotaping lessons, which gave us the possibility for videotaping our staged videos (see Behling et al., 2019). Each workshop day began with handing out the scripts to the students and determining the speaking students in our video case. The teacher and all speaking students used microphones for improving the audio-recording quality. After text learning and practicing, video cases were recorded, in which both the teacher and the students were visible (cf. Bliss & Reynolds, 2004). The selected perspective of the video situations was comparable to the perspective of an observing person sitting in the back of the classroom and monitoring activities in the classroom. The average length of the relatively self-contained video cases was 2.56 min ($SD = 2.08; Min = 67; Max = 5.23$).

4.2. Sample size
We contacted German secondary schools to recruit participating teachers, which completed university education and their teacher trainee program. Once the theoretical and practical training had been completed, teachers were expected to have the necessary knowledge and skills to plan, design, and reflect on lessons according to scientific standards. Therefore, we assumed that, while watching the videos in DiKoBi Assess, these in-service teachers must be able to identify the challenging classroom situations we had scripted in the video cases (cf. Piwowar et al., 2018). After the recruitment process, five teachers agreed to participate in the study. This qualitative approach follows other studies in which constructs such as PCK could be validly recorded with eleven or less participants (e.g., Alonzo & Kim, 2016; Bertram & Loughran, 2012; Jüttner & Neuhaus, 2013; Sorge et al., 2019).

For validation, we interviewed the teachers using think-aloud protocols (see Table 3). The interviews were videotaped. All teachers taught the subject biology in German secondary schools (with
Table 2. Overview about subject-specific dimensions of instructional quality, assigned classroom situations, and challenging aspects of biology instruction in the simulated classroom situations. Participants have to refer to these challenges when watching the videos

| Subject-specific dimension | Classroom situation | Challenging aspects |
|---------------------------|---------------------|---------------------|
| Level of students’ cognitive activities and creation of situational interest | (1) Activation of prior knowledge and ideas | Level of students’ cognitive activities. Lesson starts with a simple nomination of the topic. Cognitively stimulating components like a cognitive conflict or a focus question are missing. |
| Dealing with students’ ideas and errors | (2) Dealing with students’ ideas and errors | Formative handling of students’ errors. Students’ errors are rarely noted. The teacher provides no feedback. No effort to support conceptual change. |
| Use of technical language | (3) Use of technical language | Quantity and quality of technical terms. Teacher uses a high number of technical terms. Not all of them are appropriate for school learning. |
| Use of experiments | (4) Use of experiments | Scientific inquiry. Experiment is only used for illustration. No elements of scientific reasoning are considered. |
| Use of models | (5) Use of models | Elaborate model use. Models are only used for illustration. The application for scientific inquiry is missing. |
| Conceptual understanding | (6) Summary/transfer of knowledge | Referring back to the beginning of the lesson. At the end of the lesson, there is no referring back to the initial conflict/focus question/assumptions. |

Grade 5 to 12). The interviewed teachers had an average age of 40.4 years (SD = 9.2) and an average teaching experience of 9.4 years (SD = 6.9) after their teacher trainee program.

4.3. Description of the validation process

As DiKoBi Assess is intended to measure aspects of teachers’ professional vision when they are noticing and reasoning about biology instruction, we follow a construct-centered approach (Messick, 1994b). We examined the validity of the content of the videos (RQ1, RQ2) (hereinafter referred to as “Part A”), and that of the assigned tasks (RQ3) (hereinafter referred to as “Part B”) by conducting interviews using think-aloud protocols. Using think-aloud protocols in interviews is considered as an appropriate method in the development of instruments and items, as they provide insight into the thoughts and views of teachers (Hill et al., 2007).
Videotaped interviews using think-aloud protocols within this study were conducted with the participating teachers in April 2018. Lasting for an average of 120 minutes, each interview began with an introduction and an exercise for training to use the method thinking aloud (see Figure 2). Next, the five participating teachers watched a video showing a classroom situation of another topic of biology instruction and thought aloud how they would assess the classroom situation. After this training, they were informed about the aim of the study, the implementation of the interviews, and the tasks to be executed. Furthermore, the topic of the given classroom situation was linked to the curriculum using a sequence schedule. Then, Parts A and B of the validation using think-aloud protocols started as the main study.

Table 3. Background information on the five interviewed biology teachers

| Name | Paul | Jessica | Carla | Tom  | Ben  |
|------|------|---------|-------|------|------|
| Gender | Male | Female | Female | Male | Male |
| Age   | 39   | 31      | 53    | 33   | 46   |
| Education | Second State Teacher examination | Second State Teacher examination | Second State Teacher examination | Second State Teacher examination |
| Number of taught biology lessons per week (hours) | 12 | 9 | 15 | 8 | 8 |
| Teaching experience (years) | 7 | 1 | 17 | 6 | 16 |
| Taught subject besides biology | Chemistry | Chemistry | Chemistry | Chemistry | Chemistry |

Note. Teachers’ names were changed to pseudonyms.

Figure 2. Steps of the validation shown as a flow diagram for analysis of Part A. First, based on empirical studies subject-specific features of instructional quality were described, and a category system was derived. Subsequently, the staged videos were recorded. Then interviews using think-aloud protocols with in-service teachers have been conducted. After the transcription of the responses to these protocols, the category system was used for coding and analyzing the transcripts.
To illustrate how teachers thought aloud, and how we analyzed Parts A and B of the interviews, the following sections give an overview about the category systems used for coding, examples of the coding process, and a description of the empirical analysis of the data for each part.

4.3.1. Part A: validation of content of videos
First, in-service teachers watched the staged videos showing the six classroom situations. When they identified challenging aspects about teaching and learning, they stopped the video and talked about their thoughts aloud. Each time when teachers stopped the video, the time was noted on the log sheet by the interviewer.

4.3.1.1. Assessing authenticity. To assess the extent to which each staged classroom situation was perceived as authentic, teachers rated three follow-up questions about the authenticity of the watched classroom situations with five-point Likert-scale response options (see Table 4; Schubert et al., 2001; Seidel et al., 2010). When teachers rated the items low, they had to justify their answer.

4.3.1.2. Category system. The category system for coding Part A of the interview using think-aloud protocols was derived from research on biology-specific instructional quality features, and included challenging aspects of biology instruction, which can be found in the recorded classroom situations (see Figure 2). The categories of the category system are identical with the challenges listed in Table 2. In addition, we added the categories further PCK-aspects and not relevant. The category further PCK-aspects was added to each classroom situation for coding in case teachers mentioned other challenges of the specific situation.

The categories of the classroom situations used for coding are therefore as follows:

- (1) Activation of prior knowledge and ideas:
  - level of students’ cognitive activities,
  - creation of situational interest,
  - further PCK-aspects
- (2) Dealing with students’ ideas and errors:
  - formative handling of students’ errors,
  - student error 1,
  - student error 2,
  - further PCK-aspects
- (3) Use of technical language:
  - quantity of technical terms
  - quality of technical terms
  - explaining terms and linguistic understanding
  - further PCK-aspects
- (4) Use of experiments:
  - scientific inquiry
  - further PCK-aspects
- (5) Use of models:
  - elaborate model use
  - critical reflection
  - further PCK-aspects
Table 4. Description of the scale authenticity

| Item | (1) Activation of prior knowledge and ideas | (2) Dealing with students’ ideas and errors | (3) Use of technical language | (4) Use of experiments | (5) Use of models | (6) Summary/ transfer of knowledge |
|------|-------------------------------------------|-------------------------------------------|----------------------------|----------------------|-----------------|----------------------------------|
| Item 1 | I appraise the classroom situation as authentic. |                                           |                            |                      |                 |                                  |
| Item 2 | The teaching situation seemed to be just like a real classroom situation. |                                           |                            |                      |                 |                                  |
| Item 3 | The experience of the teaching situation in the video was similar to the experience of a real classroom situation. |                                           |                            |                      |                 |                                  |

Note. In-service teachers rated the three listed items with Likert-scale response from 1 (disagree) to 5 (agree) for each classroom situation.
(6) Summary/transfer of knowledge:
  o referring back to the beginning of the lesson
  o conceptual understanding
  o further PCK-aspects

4.3.1.3. Coding procedure. We analyzed the transcripts through qualitative content analysis (Mayring, 2014). Thus, we first divided the transcripts into coding units. For validation of the content of the videos, one coding unit of the transcript corresponded to one content-specific analytic proposition (N = 397 coding units). As we wanted to analyze if the teachers could identify challenging classroom situations in the staged videos, we defined each identified challenging aspect as one coding unit. After dividing the transcripts into coding units, we analyzed the material using a structuring analysis technique (Mayring, 2014). All structuring was done using MAXQDA 2018 (VERBI Software, 2017). For analysis of the transcripts, we coded each coding unit of the transcripts with the category system. In case teachers identified challenges other than the ones we staged, we discussed the actual visibility of the perceived challenge in the classroom situation, and added appropriate challenging aspects as examples of the category further PCK-aspects in the coding manual. The steps of the validation process are graphically shown for Part A in Figure 2.

The following excerpt of the transcript of Teacher Paul’s interview illustrates how the teacher thought aloud when he worked on the first staged video clip, as well as the coded categories from the category system (see Table 5).

4.3.1.4. Empirical analysis. After we had coded all identified challenging aspects by means of the category system, we checked each teacher’s thoughts and each classroom situation to find out whether or not our staged challenging classroom situations were identified. Eventually, we summed up the numbers of identification. For later evaluation of the data, it was important that each interviewed teacher referred to at least one challenging aspect of each of the classroom situations. The staged challenges thus served as indicators of the classroom situations, which

| Table 5. Excerpt of the transcript of Teacher Paul’s interview using the think-aloud protocol when he worked on the first staged video clip |
|---|---|
| Transcript | Coded category from the category system |
| [Paul watched video 1, and then stopped the video] So, the teacher tried to relate the new topic to the one of the previous lesson by activating the student’s prior knowledge, but I couldn’t make out a proper introductory stage at the beginning of the lesson. She did not present an interesting problem that would be the foundation of the lesson, nor did she encourage her students to formulate a leading question. | Identified challenging aspect (cognitive activation) |
| [Paul continued watching video 1, and then stopped again.] It is not clear at the moment, whether she [the teacher in the video] tries to promote the student’s curiosity by confronting them with a cognitive conflict or, for example, with an issue close to their everyday life. Basically, she briefly related the new topic to the old one. Then, as soon as one of the students dropped the word, she revealed to the class that, yes, indeed, the new topic is going to be skin. There is nothing that sparks curiosity; nothing that may carry the class through the lesson. [...] | Identified challenging aspects (creation of situational interest and cognitive activation) |

Note. Excerpt translated from German by a bilingual speaker.
focused on different biology-specific dimensions of instructional quality, whereby the identification of one indicator was sufficient for a positive coding of the challenging classroom situation.

4.3.2. Part B: validation of assigned tasks
After watching the six staged videos of the first lesson, the teachers worked on the first classroom situation (Activation of prior knowledge and ideas) in DiKoBi Assess. Therein, teachers worked on the tasks in the assessment tool while simultaneously thinking aloud (Aitken & Mardegan, 2000; Ericsson, 2006). As the three tasks Describe, Explain, and Alternative Strategy are the same for every classroom situation, we focused on the first classroom situation for the construct validation as an example.

4.3.2.1. Category system. The category system for validating the assigned tasks was derived from the theory of professional vision, and included a description of the aspects of professional vision and coding examples (Seidel & Stürmer, 2014; Van Es & Sherin, 2002; see Table 6).

First, the teachers watched the video of the first classroom situation Activation of prior knowledge and ideas and identified challenging aspects by making notes about their identified aspects in open text fields on the computer (Task Describe). We assumed Task Describe to measure description as an aspect of professional vision. Second, participants reasoned about their described challenges by linking them to scientific theories and concepts (Task Explain). Additionally, they estimated their confidence about their reflections on the classroom situation by adjusting a 10-point-slider on a questionnaire scale (between “completely unconfident” to “very confident”). We assumed Task Explain to measure explanation as an aspect of professional vision. Third, the participants proposed an alternative teaching strategy, and justified their selected strategy regarding the question why it would improve the classroom situation (Task Alternative Strategy). Again, for each classroom situation they had to estimate their confidence.

| Task                | Examples of statements | Aspect of professional vision                  | Description from the coding manual                                                                 |
|---------------------|------------------------|------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| Describe            | [1] He just presents the topic. | Description                                     | Specific and differentiated observation of the challenging aspect without reasoning, evaluations or judgements. |
| Explain             | [2] there is nothing that functions as a catch. | Explanation                                     | Answer is a rationale including a theoretical reference, which is made using PCK/PK theories, concepts, and technical terms. |
|                     | [3] Therefore the students are not taken to the topic of the lesson. | Prediction as a learning consequence             | Effects or consequences of the observed events for the learning/learning development of the students are derived. |
| Alternative Strategy| [4] During a productive stage of the lesson one could compare skin and its functions to a jacket and its functions, for example. | Prediction as an alternative strategy            | Suggestions of alternative responses to student answers or alternative teaching strategies that can make teaching more effective. |

Note. Statements translated from German by bilingual speaker.
about their described and explained teaching alternatives. For Task Alternative Strategy, we assumed predicting as giving an alternative teaching strategy as an aspect of professional vision.

Table 6 shows excerpts from Part B of the transcripts as well as the coded categories, which are the aspects of professional vision. Additionally, the description about how the aspects of professional vision are defined and operationalized in the coding manual is given. For Task Explain, two subcategories were defined to subdivide the coding rubrics more finely, and to consider explanation and prediction as aspects of professional vision as described by Seidel and Stürmer (2014). The subcategories could either contain rationales that referred to a specific concept or contained a conceptual component as a technical term (explanation), or statements that touched on conceivable consequences of the observed events (prediction as a learning consequence). However, intended answers to task Explain should focus on the linking between observation and appropriate theories that were coded with the professional vision aspect explanation.

4.3.2.2. Coding procedure. Part B of each transcript was cut into coding units on the statement level (N = 433 coding units). By doing so, we were able to check whether the statements reflect the assumed aspect of professional vision. For analysis of the transcripts, we coded each single statement with the category system of professional vision.

4.3.2.3. Empirical analysis. After coding all statements, the codings were aggregated per task. Thus, we summed up the codings of each task separately and set the sum of each coded category in relation to the total number of statements for the corresponding task.

4.4. Interrater reliability
For reliability and validity assurance, two independent raters analyzed Parts A and B of one transcript with the category systems. The raters were research assistants, who first familiarized themselves with the different category systems and the connected concepts. Afterwards they were trained in coding by taking another transcript for practicing. The training included several coding and discussion loops within a period of about one month. For assessing interrater reliability, 124 coding units were analyzed for Part A, and 265 coding units for Part B. Calculation of Cohens Kappa showed very good agreement: \( \kappa = .86 \) for Part A; \( \kappa = .95 \) for Part B; \( \kappa = .93 \) in total (Wirtz & Caspar, 2002).

5. Results
We summarize the results here regarding the research questions of Parts A and B. Part A refers to the validation of the content of the videos, and shows the results of teachers’ self-reports on authenticity (RQ 1), and their identified challenges of the classroom situations (RQ 2). Part B refers to the validation of the assigned tasks, and to the results of teachers’ use of the aspects of professional vision (RQ 3).

5.1. Part A: validation of content of videos
To assess authenticity of the simulated classroom situations, the teachers answered follow-up questions after watching each simulated video. Table 7 shows the results of the teachers’ self-reports on authenticity of the classroom situations. All of the authenticity ratings show high values. As reported by the teachers, none of the classroom situations seemed to be artificial or unnatural.

Table 8 shows how often the challenging classroom situations were identified by the teachers. For a deeper look into the teachers’ identifications, we also counted each identified challenging aspect that served as indicator of the challenging classroom situations.

The teachers identified almost every challenging classroom situation, though with varying frequency. A lack of problem identification occurred in the second classroom situation that focused on teachers’ dealing with students’ ideas and errors. Here, we embedded three different challenging aspects in the situation. The first identifiable challenging aspect referred to the observation
Table 7. Descriptive statistics for authenticity of the classroom situations

| Item   | (1) Activation of prior knowledge and ideas | (2) Dealing with students’ ideas and errors | (3) Use of technical language | (4) Use of experiments | (5) Use of models | (6) Summary/transfer of knowledge |
|--------|--------------------------------------------|---------------------------------------------|-------------------------------|-----------------------|------------------|----------------------------------|
| Item 1 | M (SD) 4.80 (.45)                           | M (SD) 4.80 (.45)                           | M (SD) 4.40 (.89)             | M (SD) 5.00 (.00)    | M (SD) 5.00 (.00) | M (SD) 5.00 (.00)                |
| Item 2 | M (SD) 5.00 (1.00)                          | M (SD) 4.80 (.45)                           | M (SD) 4.60 (.55)             | M (SD) 5.00 (.00)    | M (SD) 5.00 (.00) | M (SD) 5.00 (.00)                |
| Item 3 | M (SD) 4.40 (.89)                           | M (SD) 4.80 (.45)                           | M (SD) 4.20 (.84)             | M (SD) 5.00 (.00)    | M (SD) 5.00 (.00) | M (SD) 5.00 (.00)                |
| Total scale | M (SD) 4.73 (.59)                        | M (SD) 4.80 (.41)                           | M (SD) 4.40 (.74)             | M (SD) 5.00 (.00)    | M (SD) 5.00 (.00) | M (SD) 5.00 (.00)                |

Note. Likert-scale response options from 1 (disagree) to 5 (agree), answered by the five in-service teachers. For each classroom situation, the mean and standard deviation of the individual items of the follow-up questions as well as the mean of the authenticity scale are shown.
that the teacher does not use students’ answers and errors’ in a formative and supportive way, but ignores them mostly. Furthermore, the teacher in the video does not notice wrong statements given by two students. We listed the wrong statements as another two indicators (student’s error 1 and student’s error 2). An example of Teacher Ben’s statement to identify this challenge for which he was given a positive coding is: “Well, er, of course, this [teachers’ behavior in the video] is problematic, because the student gave a completely different answer. And the teacher, well, she has her own idea of what should be written down”. Ben referred to the teacher’s inappropriate handling of the answer given by the student, because the teacher in the video did not deal with the original answer of the student. Apart from Ben, none of the other teachers mentioned this behavior of the teacher in the video clip.

Even if the other numbers of identification were acceptable, a deeper look at the individual challenges revealed another lack of identification within classroom situation (6). One problem that should be identified related to the teacher’s missing reference back to the beginning of the lesson. Teacher Clara said that “the teacher should go back to her mind map [from the beginning]. I think it did not help the students to memorize the results of the lesson, the way she did it”. However, none of the other teachers mentioned this challenge.

Additionally, the teachers identified further biology-specific challenges in the classroom situations other than the scripted ones. Further challenges they identified mostly referred to different aspects of cognitive activation, complexity of instruction, dealing with students’ answers, and scientific working methods. Apart from that, they mentioned general pedagogical issues such as the use of media, social arrangements, general teaching methods, time management, classroom management, achievement of learning goals, and the general structuring of the lesson.
5.2. Part B: validation of assigned tasks

To investigate whether the tasks in DiKoBi Assess measure the assumed aspects of the construct of professional vision, we assigned to each of the tasks (Describe, Explain, and Alternative Strategy) aspects of professional vision, which occurred while teachers used DiKoBi Assess. For analysis, 219 relevant coding units were identified. When working on Task Describe, teachers mostly used description (57 % of the coding units) and alternative strategies (37 % of the coding units). Teachers’ answers regarding Task Explain could mostly be assigned to explanation (88 % of the coding units), whereas those regarding Task Alternative Strategy could be assigned to prediction as alternative strategy (100 % of the coding units) (see Figure 3).

6. Discussion

Regarding our objectives, we wanted to develop an instrument for measuring aspects of professional vision of biology teachers to approximate their ePCK_{reflect}. Therefore, we developed DiKoBi Assess that represents real-world biology classroom situations. Through the implementation of instructional tasks that guided in-service teachers’ professional vision, combined with the request to think-aloud, we believe that our approach enabled us to infer teachers’ ePCK_{reflect}. The participating teachers had to directly articulate their thoughts on particular classroom situations and on the performance of the teacher in the video, and thus, had to draw on a specific subset of knowledge and skills while doing these assessments (cf. Alonzo et al., 2019). Therefore, we claim that using professional vision as a measure to approximate forms of PCK opens promising possibilities.

This study addressed three research questions, which are discussed below.

6.1. Part A: validation of content of videos

When validating the content of the videos, we wanted to find out whether the interviewed in-service teachers assessed the simulated classroom situations as authentic (RQ1), and whether the teachers were able to identify the staged challenging classroom situations (RQ2). The results of this study revealed that the staged videos of the six classroom situations were perceived as authentic representations of practice, thus supporting present results from research on videos capturing the authenticity of classroom practice (Codreanu et al., 2020; Piwowar et al., 2018). Besides the inclusion of in-service teachers in the validation process to strengthen ecological validity, we made several efforts to achieve authenticity of the staged classroom situations. For example, we have based our conceptual design on the national curriculum and on previous research results on instructional quality in order to define self-contained challenging aspects for each classroom situation, thus ensuring an explicit focus (cf. Bliss & Reynolds, 2004; Piwowar et al., 2018). The use of video situations in which features of instructional quality were not perfectly implemented may have additionally contributed to an authentic perception of the classroom situations. Other possible authenticity enhancing components are the chosen camera perspective showing both the teacher and students (thus imitating the perspective of a person sitting in on classes), the relatively short length of the videos not exceeding 5.23 min, and providing contextual
information (cf. Bliss & Reynolds, 2004; Gold & Holodynski, 2017; Müller, 2016). However, our present results do not provide any information about the level of presence or the degree of immersion, which can also be regarded as indicators of authenticity, and thus, need to be evaluated in subsequent studies (Codreanu et al., 2020; Schubert et al., 2001).

Regarding the second research question, the results of content validation showed that almost all of the challenging classroom situations were identified sufficiently. Individual challenging aspects served as indicators of the classroom situations so that the participants who were using the assessment tool did not have to perceive two challenges at the same time. Thus, a positive coding could already be reached by referring to one of the challenging aspects. We justify this decision on coding by two reasons. First, we took into account the difficulty of relating conceptual professional knowledge to specific practical assessment situations in classrooms. Our decision followed the idea of the decomposition of practice into aspects of classroom situations that support teachers in their professional development (Grossman et al., 2009). Second, since we wanted to measure aspects of professional vision with our instrument, we focused on teachers’ abilities to describe, explain, and suggest alternative strategies, but we did not require them to name all indicators of a classroom situation.

Nevertheless, the teachers’ varying abilities to identify the challenging aspects also depends on their knowledge, beliefs, and prior experiences that might change the way they noticed complex situations (Borko, 2004; Van Es & Sherin, 2008). Although we did not systematically evaluate the responses of the participating teachers in this study, we would like to emphasize that specific experiences led to teachers’ more differentiated abilities to notice and reason in the classroom. From our observations, this was particularly the case for teachers who, in addition to their own teaching practice, were also engaged in the education of future teachers (e.g., as seminar instructors in teacher training). At the same time, due to the small number of participants in this study, such observations cannot be generalized.

However, the result of teachers’ very poor identification of challenging aspects within the second and the last classroom situations called for instructional changes to improve the instrument as a whole. The second classroom situation referred to teachers’ dealing with students’ errors. For successful problem identification, the teachers had to focus on student activities to find out students’ errors first. This was an essential condition to further judge the situation. In previous studies, researchers emphasized that in-service teachers in particular focus on teachers’ behavior in videos rather than on the activities of their students (Colestock & Sherin, 2009; Gaudin & Chaliès, 2015; Santagata, 2009; Van Es & Sherin, 2008). Our results confirmed this observation. Instead of focusing on the students’ answers, the participating teachers in our study focused on teachers’ behavior. For example, the teachers mentioned a low level of students’ cognitive activities, the complexity of the given instructions, and missing methods for visualization. All of these examples of criticism result from teachers’ actions in the classroom situation. Hence, we decided to adjust the instructional task to reduce construct-irrelevant variance by adding an observation focus on the handling of students’ responses (Messick, 1994a). Another in-service teacher’s response to an additional think-aloud protocol showed that the scripted challenge could be identified clearly after task modification. Furthermore, one of the challenges included in the last classroom situation, in which the acting teacher in the video did not refer back to the start of the lesson, was also not identified satisfactorily by the participating biology teachers. We assume that this aspect was not obvious to them to criticize, because there was neither a specific part during the start of the lesson like a focus question which needs to be answered, nor a problem which needs to be solved, nor a context to which the teacher in the video can refer back. Thus, we inserted another video clip before the sixth classroom situation. The alternative teaching strategy illustrates how to introduce the topic skin and its functions by using an analogy of a knight’s armor. During the lesson, students have to elaborate essential outcomes regarding the topic. After summarizing the main points, the teacher should refer back to the analogy used at the beginning of the lesson. Thus, the acquired knowledge of the topic skin can be transferred to teaching other content areas so that content linking, and therefore knowledge linking, can be supported (Förtsch et al., 2020;
Scott et al., 2011). When the teacher refers back to a focus question, a problem or a context, students have to use their gained knowledge to explain the underlying relations. In doing so, unconnected knowledge elements are activated and connected. This way of science teaching requires the consideration of the conceptual change theory that is a fundamental theoretical perspective of how science learning takes place (Nachreiner et al., 2015; Özdemir & Clark, 2007).

Besides our scripted challenges, the participating biology teachers noticed further biology-specific problems that they mostly referred to as different aspects of cognitive activation, complexity of instruction, dealing with students' answers, scientific working methods, and general pedagogical issues. When using videos of classroom situations, it is not possible to show a perfect teaching situation as there are too many influencing variables (Seidel & Stürmer, 2014). Among all the occurring events of teaching and learning, teachers have to notice those events with particular importance for students' science learning within our video-based assessment tool. This ability to notice events that are relevant to students' science learning is the professional competence of teachers' subject-specific vision (Seidel & Stürmer, 2014; Steffensky et al., 2015).

6.2. Part B: validation of assigned tasks

In response to RQ3, we examined whether the tasks measured the aspects of the conceptualizations of teachers' professional vision. Overall, teachers' answers to the tasks fitted our assumptions for most parts. The intended aspects describing, explaining, and predicting were measured during reasoning.

Regarding Task Describe, we asked teachers for a brief description of challenging aspects to measure their professional vision ability of describing. The results showed that the interviewed teachers used both description and prediction as alternative strategy for answering this task. It was not our expectation that teachers would use alternative strategies to answer Task Describe. However, as in-service teachers possess more knowledge about classroom practice than pre-service teachers, the strategy to reason by explaining alternative teaching performance might be part of this expertise (Jong et al., 2012; Palmer et al., 2005). Additionally, knowledge can occur as encapsulated knowledge (Boshuizen & Schmidt, 1992; Schmidt & Rikers, 2007) that can be described as “existing declarative-conceptual knowledge [...] enriched with experience from cases” (Heitzmann, 2014, p. 28). Thus, this experience may be used for evaluating effective teaching. However, a close look at the teachers' answers also showed that the description of the challenge was indirectly present in their answers. For example, when a teacher said that it would be better to choose an introduction that is close to everyday life, the reverse conclusion is that the teacher in the simulated video did not use an introduction that was close to everyday life. Therefore, we decided to expand our coding manual and include examples of indirect descriptions as alternative answers of Task Describe. However, it would also be necessary to analyze whether the same answering behavior occurred among pre-service teachers, or whether the observed behavior in our study was based on the practical experience of the in-service teachers.

Teachers' answers to Task Explain were mainly coded as explanation. This result is in line with our expectation of measuring the professional vision aspect explanation. For positive codings, participating teachers had to reason about the observed classroom specifics by connecting them with professional terms and concepts (Seidel & Stürmer, 2014). However, the rationales could theoretically also refer to consequences for students and their learning (prediction as learning consequence), even if this was not explicitly required in the task definition. Since this type of a rationale was rare and not obviously detectable due to the similarity of the given rationales, we decided to combine the subcategories into just one coding category for future coding. Regarding teachers' answers of Task Alternative Strategy, we measured the situation-specific skill prediction. All answers could be categorized as giving alternative teaching strategies that benefit the acquisition of professional decision making (Kersting et al., 2010; Van Es & Sherin, 2002). In light of a construct-centered approach (Messick, 1994b), we can summarize that our assigned tasks are suitable for measuring the aspects of biology-specific professional vision.
6.3. Conclusions and outlook

Our validation study showed that DiKoBi Assess can be used as a valid test instrument for measuring professional vision of biology teachers. It extends the basic dimensions for science education that currently used instruments can measure when investigating teachers’ professional development in a subject-specific environment. To complete the validation process, data of pre-service teachers and in-service teachers are currently being compared. Initial analyses indicated that in-service teachers may achieve better scores in the area of professional knowledge based on their teaching experience.

In addition to the validation of the instrument, the interviews using think-aloud protocols showed that DiKoBi Assess triggered both teachers’ reflection on their perceptions of good or bad practice and their reflection on whether their perceptions were based on a theoretical basis or only on their teaching experience. A glimpse of this conclusion can be seen in the answer of teacher Ben, who talked about his reflection: “(you start thinking about your judgment), whether it is just a gut feeling or whether there is more behind it”. For reflection on teaching strategies, empirical evidence and theories, teachers need to draw on their ePCK_{reflect} (Alonso et al., 2019). Thus, reflection fosters teachers’ professional development, and thus, increases their professional competence (Blomberg et al., 2013; Gaudin & Chaliès, 2015; Santagata & Yeh, 2014). Learning becomes even more effective when video clips are used with guidance from facilitators, who help learners to identify and make sense of relevant content in video clips. This is a promising approach in actual teacher education programs (Blomberg et al., 2013; Van Es et al., 2014). However, the knowledge teachers likely apply in the moment of classroom instruction is considered to be important for high-quality instruction (Kersting et al., 2012). Thus, using DiKoBi Assess as a constructive alignment (Biggs & Tang, 2007) of biology classroom situations activates teachers’ subject-specific knowledge in the context of biology education and supports the establishment of learning opportunities for students.

In future studies, DiKoBi should not only be used for measuring, but also for fostering professional vision of biology teachers (DiKoBi Learn). Hence, the suitability for fostering aspects of professional vision needs to be examined systematically.

Funding
This research was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG) to the project Facilitating Diagnostic Competences in Simulation-based Learning Environments in Higher Education (COSIMA; grant number NE 1196/8-1, FOR 2385).

Author details
Maria Kramer
E-mail: maria.kramer@bio.lmu.de
ORCID: http://orcid.org/0000-0003-4243-2867
Christian Förtsch
E-mail: christian.foertsch@bio.lmu.de
ORCID: http://orcid.org/0000-0002-3625-6791
Julia Stürmer
E-mail: julia.schauberger@bio.lmu.de
Sonja Förtsch
E-mail: sonja.foertsch@biologie.uni-muenchen.de
Tina Seidel
E-mail: tina.seidel@tum.de
Birgit J. Neuhaus
E-mail: didaktik.biologie@lrz.uni-muenchen.de
1 Biology Education, Department 1, Faculty of Biology, Ludwig-Maximilians-Universität München, Munich, Germany.
2 Friedli Scholler Endowed Chair for Educational Psychology, School of Education, Technical University of Munich (TUM), Germany.

References
Altken, L. M., & Mardegan, K. J. (2000). Thinking aloud: data collection in the natural setting. Western Journal of Nursing Research, 221, 841–853. https://doi.org/10.1777/019394500022064-791
Alonso, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers’ PCK in action: Enacted and personal PCK. In A. Humre, R. Cooper, & A. Borowski (Eds.), Repositioning pedagogical content knowledge in teachers’ knowledge for teaching science (pp. 271–286). Springer.
Alonso, 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
Barab, S. A., Squire, K. D., & Duerer, W. (2009). A co-evolutionary model for supporting the emergence of authenticity. Educational Technology Research and Development, 48(2), 37–62. https://doi.org/10.1007/BF02313400
Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. Cognitive Science, 16(2), 153–184. https://doi.org/10.1012/jteo.20255

Brunner, M., Kunter, M., Krauss, S., Baumert, J., Voss Blum, W., Dubberke, T., Jordan, A., Loewen, K., & Tsai, Y.-M. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COACTIV-Projekts [Mathematics Teachers’ Professional Competence: Conceptualization, Measuring, and Importance for Teaching: Interim Results of the COACTIV Project]. In M. Prezel & L. Allolio-Näcke (Eds.), Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG Schwerpunktprogramms [Studies on Educational Quality: Final Report of the DFG Priority Program] (pp. 54–82). Waxmann.

Carlson, J., & Kaiser, K. R. (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. 77–92). Springer.

Cordero, E., Sommerhof, 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, 103146. https://doi.org/10.1016/j.tate.2020.103146

Colestock, A., & Sherin, M. G. (2009). Teachers’ sense-making strategies while watching video of mathematics instruction. Journal of Technology and Teacher Education, 17(1), 7–29.

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

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

Dorffner, T., Förtsch, C., Gern, 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

Dorffner, T., Förtsch, C., & Neuhaus, B. J. (2017). Die methodische und inhaltliche Ausrichtung quantitativ-einder Videozuständigkeiten der 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. A Review). Zeitschrift für Didaktik der Naturwissenschaften, 23(1), 261–285. https://doi.org/10.1007/s40573-017-0058-3

Dorffner, 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
Dorfer, T., Forstch, 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/10.1080/02635434.2019.1609436

Dorfer, T., Forstch, C., Spangler, M., & Neuhaus, B. J. (2019). Wie plane ich eine konzeptorientierte Biologiestunde? Ein Planungsmodell für den Biologieunterricht - Das Schalenmodell. - Der mathematische und naturwissenschaftliche Unterricht (MMU), 4, 300–306. http://www.didaktik.bio.lmu.de/download/aktuelles/2019/mmU_04_2019_300-306.pdf

Ericsson, K. (2006). Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance on Representative Tasks. In K. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), The Cambridge Handbook of Expertise and Expert Performance (Cambridge Handbooks in psychology). pp. 223–242. Cambridge University Press. https://doi.org/10.1017/CBO9780511811679.013

Fischer, H. E., Borowski, A., & Tepner, O. (2012). Professional knowledge of science teachers. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), Springer International Handbooks of education: Vol. 24.2. Second International Handbook of science education (pp. 435–448). Springer.

Forstch, C., Dorfer, T., Baumgartner, J., Werner, S., Kotzebue, L., & Neuhaus, B. J. (2020). Fostering students' conceptual knowledge in biology in the context of German National Education standards. Research in Science Education, 50(2), 739–771. https://doi.org/10.1080/11165018.018.9709-8

Forstch, C., Heidenfelder, K., Spangler, M., & Neuhaus, B. J. (2018). How does the use of core ideas in biology lessons influence students' knowledge development? Zeitschrift für Didaktik der Naturwissenschaften, 44 (1), 271. https://doi.org/10.15473/sdzn.04573.018-0071-1

Forstch, C., Werner, S., Dorfer, T., Kotzebue, L., & 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

Forstch, C., Werner, S., Kotzebue, L., & 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

Gaudin, C., & Chalès, 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

Gold, B., & Holodnyński, 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. https://doi.org/10.1525/aa.1994.96.3.02.000100

Grossman, P., Compton, C., Jira, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. Teachers College Record, 111, 2055–2100. https://doi.org/10.2307/j.ctv4z86k

Großmann, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. American Educational Research Journal, 45(1), 184–205. https://doi.org/10.3102/0020311306182155

Hattie, J. (2009). Visible learning: A synthesis of meta-analyses relating to achievement. Routledge.

Heidenfelder, K. (2016). Strukturierung des Biologieunterrichts mit Basiskonzepten in Verbindung mit problemorientierten Kontexten [Structure of biology lessons with core ideas and problem-oriented contexts] [Dissertation]. Ludwig-Maximilians-Universität München.

Heitzmann, N. (2014). Fostering diagnostic competence in different domains [Dissertation]. Ludwig-Maximilians-Universität München.

Herppich, S., Wittwer, J., Nückles, M., & Renkl, A. (2013). Benefits for processes cause decrements in outcomes: Training improves tutors’ interaction at the expense of assessment accuracy. In M. Knauff, M. Pouen, N. Sebanc, & J. 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

Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. Measuring Disclosure in Interdisciplinary Research and Perspectives, 5(2–3), 81–92. https://doi.org/10.1080/15366360701468999

Hoth, J., Kaiser, G., Döhrmann, M., König, J., & Blommeke, S. (2018). A situated approach to assess teachers' professional competencies using classroom videos. In O. Buchbinder & S. Kunze (Eds.), Mathematics teachers engaging with representations of practice: A dynamically evolving field (pp. 23–45). ICME-13 Monographs, Springer.

Jong, J. R. J., De, van Tartwijk, J., Verloop, N., Veldman, I., & Wubbels, T. (2012). Teachers’ expectations of teacher-student interaction: Complementary and distinctive expectancy patterns. Teaching and Teacher Education, 28(7), 948–956. https://doi.org/10.1016/j.tate.2012.04.009

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/10.11141/jets.v1i2.126

Kane, T. J., McCaffrey, D. F., Miller, T., & Stoger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. Bill & Melinda Gates foundation. https://fileseric.ed.gov/fulltext/ED540959.pdf

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., 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

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. https://doi.org/10.3102/0020311312437853

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
Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. Learning and Instruction, 17(5), 494–509. https://doi.org/10.1016/j.learninstruc.2007.09.002
Kyräidies, K., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. Teaching and Teacher Education, 36, 143–152. https://doi.org/10.1016/j.tate.2013.07.010
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/10.1080/09500690802178058
Lehane, L., & Bertram, A. (2016). Getting to the CoRe of it: A review of a specific PCK conceptual lens in science educational research. Educação Química, 27(1), 52–58. https://doi.org/10.1016/j.eq.2015.09.004
Lipowsky, F., Roczkcy, K., Pauli, C., Drollering-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-comings: Students’ understanding of the pythagorean theorem. Learning and Instruction, 19(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001
Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. Journal of Research in Science Teaching, 41(4), 370–391. https://doi.org/10.1002/tea.20007
Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlosen [Scientific Inquiry as Scientific Problem Solving]. In D. Kruger & H. Vogt (Eds.), Theorien in der biologiedidaktischen Forschung [Theories in research of biology education] (pp. 177–186). Springer.
Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution. Beltz. https://www.psychopen.eu/fileadmin/user_upload/books/mayring/ssoor-2014-mayring- Qualitative_content_analysis_theoretical_founda pdf
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
Messick, S. (1994a). Alternative modes of assessment, uniform standards of validity. ETS Research Report Series, 1994(2), i–22. https://doi.org/10.1002/j.2333- 850X.1994.tb01634.x
Messick, S. (1994b). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23(2), 13–23. https://doi.org/10.3102/0013189X023002013
Michalsky, T. (2016). 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
Müller, R. (2016). Die Quantenphysik im Spannungsfeld zwischen Fachlichkeit, empirischer Forschung und Schulpraxis [Quantum physics in the conflict between scientific expertise, empirical research and school practice]. In C. Maurer (Ed.), Gesellschaft für Didaktik der Chemie und Physik: Vol. 36. Authentizität und Lernen – das Fach in der Fachdidaktik.}

Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Berlin 2015 (pp. 13–24). Universität Regensburg.

Nachreiner, K., Spangler, M., & Neuhaus, B. J. (2015). Begeisterung und Lernen - Basiskonzepten orientierten Unterrichts [Teaching based on biological core ideas. A theoretical foundation]. MNU (Mathematik-naturwissenschaftlicher Unterricht), 68(3), 172–177.

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

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/10.1080/09500690802187058

Özdemir, G., & Clark, D. B. (2007). An overview of conceptual change theories. Eurasia Journal of Mathematics, Science and Technology Education, 3(4), 351–361. https://doi.org/10.12973/ejmste/75614

Palmer, D. J., Stough, L. M., Burdeni, J. T. K., & Gontinec, M. (2009). Identifying teachers’ expertise: An Examination of researchers’ decision making. Educational Psychologist, 40(1), 13–25. https://doi.org/10.1207/s15326985ep4001_2

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.

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

Preterorius, 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

Questback GmbH. (2018). EFS survey (Version Winter 2018).

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, 81(1), 18–30.

Roth, K. J., Binta, 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), 3. https://doi.org/10.1186/s40594-017-0001-2

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Video-based 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

Sanctagata, R. (2009). Designing video-based professional development for mathematics teachers in low-performing schools. Journal of Teacher Education, 60(1), 38–51. https://doi.org/10.1177/ 0022487108328485

Sanctagata, 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

Santagata, R., & Yeh, C. (2014). Learning to teach mathematics and to analyze teaching effectiveness: Evidence from a video- and practice-based approach. Journal of Mathematics Teacher Education, 17(6), 491-514. https://doi.org/10.1007/s10857-013-9263-2

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

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), Handbook of motivation at school (pp. 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. H., Jüttner, M., Brandenbusch, S., Sandmann, A., & Neuhaus, B. J. (2013). Development, evaluation, and validation of a paper-and-pencil test to assess two components of biology teachers’ pedagogical content knowledge concerning the ‘cardiovascular system’. International Journal of Science and Mathematics Education, 11(6), 1389–1390. https://doi.org/10.1007/s11857-012-9384-6

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. Medical Education, 41(12), 1133–1139. https://doi.org/10.1111/j.1365-2923.2007.02915.x

Schön, D. A. (1987). Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. Jossey-Bass.

Schubert, T., Friedrich, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. Presence: Teleoperators & Virtual Environments, 10(3), 266–281. https://doi.org/10.1162/10547460175349603

Scott, P., Mortimer, E., & Amtelter, 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. In E. Klieme, D. Leutner, & M. Kenk (Eds.), Zeitschrift für Pädagogik, Beilicht: Vol. 56. Kompetenzmodellierung: Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes (pp. 296–306). Beltz.

Seidel, T., & Shavelson, R. J. (2010). 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/10.3102/0034654310341017

Seidel, T., & Stürmer, K. (2016). Modeling and measuring the structure of professional vision in preservice teachers. American Educational Research Journal, 53(4), 739–771. https://doi.org/10.3102/0002831214531321

Shavelson, R. J. (2012). Assessing business-planning competence using the Collegiate Learning Assessment as a prototype. Empirical Research in Vocational Education and Training, 4(1), 77–90. https://www.pedocs.de/volltexte/2013/8274/pdf/ERVET_2012_1_Shavelson_Assessing_business_planning_competence.pdf

Sherin, M. G. (2002). When Teaching Becomes Learning. Cognition and Instruction, 20(2), 119–150. https://doi.org/10.1207/S15326900CI2002_1

Sherin, M. G., Russ, R. S., Sherin, B. L., & Colestock, A. (2008). Professional vision in action: An exploratory study. Issues in Teacher Education, 17(2), 27–46. https://files.eric.ed.gov/fulltext/EJ831297.pdf

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

Simon, H. A. (1992). Scientific discovery as problem solving. International Studies in the Philosophy of Science, 6(1), 3–14. https://doi.org/10.1080/0269859208573403

Siy, C., & Martin, S. N. (2014). Facilitating reflexivity in preservice science teacher education using video analysis and cogenorative dialogue in field-based mathematics. EURASIA Journal of Mathematics, Science & Technology Education, 10(5), 481–508. https://doi.org/10.1027/1932-4093/a000200

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 Mistakes and Developing an Error-Friendly Culture: Conceptual Foundations and Programmatic Theses for a Pedagogical Handling of Errors]. Erwagen Wissen Ethik (Streitforum für Erwägungskultur), 19(3), 274–282.

Stahnik, R., Schueler, S., & Roesecken-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. (2018). Lehrplan PLUS. https://www.lehrplanplus.bayern.de/

Steffensky, M., & Neuhaus, B. J. (2018). Unterrichtsvision in naturwissenschaftschaftlichen Unterricht [Instructional Quality in Science]. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), Theorien in der naturwissenschaftsdidaktischen Forschung [Theories in Science Education] (Vol. 47, pp. 299–313). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56320-5_18

Steffensky, M., Carle, B., Holdynski, M., & Moller, 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/s11857-014-0002-0

Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates. Zeitschrift für Psychologie, 223(1), 54–63. https://doi.org/10.1027/2151-2604/a000200

Stürmer, K., Seidel, T., & Schäfer, S. (2013). Changes in professional vision in the context of practice. Gruppendynamik und Organisationsberatung, 44(3), 339–355. https://doi.org/10.1007/s11612-013-0216-0
Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie [Practical Activities in Physics Lessons – Results of a Video Study]. Zeitschrift für Didaktik der Naturwissenschaften, 10, 51–69. https://ftp.rz.uni-kiel.de/pub/ipn/zfdn/2004/3.Tesch_Duit_051-070.pdf

Todt, E. (1977). Motivation: Eine Einführung in Probleme, Ereignisse und Anwendungen [Motivation: An introduction to problems, events and application]. Quelle & Meyer.

Treich, F. (2016). Die Entwicklung der Professionellen Unterrichtswahrnehmung im Lehr-Lern-Labor Seminar [The Development of Professional Vision in the Student-Lab]. Julius-Maximilians-Universität Würzburg.

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

Van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers’ “learning to notice” in the context of a video club. Teaching and Teacher Education, 24(2), 244–276. https://doi.org/10.1016/j.tate.2006.11.005

Van Es, E. A., & Sherin, M. G. (2017). Bringing facilitation into view. International Journal of STEM Education, 4 (1), 32. https://doi.org/10.1186/s40594-017-0088-x

Van Es, E. A., Tunney, J., Goldsmith, L. T., & Seago, N. (2014). A Framework for the Facilitation of Teachers’ Analysis of Video. Journal of Teacher Education, 65(4), 340–356. https://doi.org/10.1177/0022017714534266

VERBI Software. (2017). MAXQDA 2018. VERBI Software. https://www.maxqda.com

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/s11076-012-9390-8

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

Wirtz, M. A., & Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität [Inter-rater agreement and Inter-rater reliability]. Hogrefe.

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