Journal of Teacher Education
2025, Vol. 76(1) 42–56

DOI: 10.1177/00224871241237497
journals.sagepub.com/home/jte

Advancing Preservice Science Teachers' Skills to Assess Student Thinking On-the-Fly Through Practice-Based Learning

Tobias Hoppe[1], Tina Seidel[2], Alexander Renkl[3], and Werner Rieß[1]

Abstract

Teachers' assessment of student thinking is both difficult to attain and essential for responsive teaching in ongoing interaction during science lessons. Principles of practice-based learning provide a basis for the design of learning environments which may equip prospective teachers for this challenging task. In an experimental study ($N = 104$), we examined the extent that the use of different media types as representations of practice (video and written cases), the number of rehearsals, and the complexity of student thinking contribute to preservice teachers' acquisition of assessment skills. Our findings indicated that participants benefited equally well from video and written cases. The number of necessary rehearsals depended on the complexity of student thinking to be assessed. This finding implies that specifics of the content to be assessed need to be taken into consideration when designing learning environments for practicing assessment skills.

Keywords

practice-based teacher education, equity, professional vision, science teacher education, preservice teacher education

Introduction

Students build knowledge based on their pre-existing ideas. To provide all students with adequate opportunities to learn science, teachers should continually assess[1] their students' thinking and make instructional decisions that are responsive (Kavanagh et al., 2020; Ruiz-Primo, 2011; Ruiz-Primo & Furtak, 2007; Taber, 2017). Within these activities, teacher judgments of student thinking are essential because teachers can only act on what they perceive, and the quality of teacher judgments determines how teachers respond (Schoenfeld, 2011). Situations conducive to the assessment of student thinking, which often occur spontaneously when interacting with the student, have been described as "teachable moments" (Shavelson et al., 2008). To provide students with adequate learning opportunities, teachers need to make note of relevant student statements and hypothesize about the underlying conceptions "on-the-fly," all in the course of a buzzing lesson (Ruiz-Primo, 2011).

Practice-based learning offers a promising approach to enable preservice teachers to acquire the skills they need to engage in such complex classroom practice (Grossman et al., 2009; Kloser, 2014; McDonald et al., 2013; Stroupe et al., 2020). Grossman (2009) suggested to identify essential components of core teaching practices. Assessing student thinking represents a component of the more complex practice of eliciting, assessing, and using student thinking for further instruction. Novice teachers should be allowed to enact such practices in complexity-reduced yet authentic learning environments. Participants then have the chance to identify and engage in elements of teaching without facing the pressure of a real classroom situation. To illustrate practice components, representations of practice have been used, for example, written and video cases (Farrell et al., 2022; Kramer et al., 2020). Although studies have shown that both media types are suitable for fostering teachers' skills in noticing and interpreting relevant aspects of classroom situations, how participants benefit from the different qualities of each media type remains empirically an open question (Friesen & Kuntze, 2018; Hoppe, Renkl, Seidel, et al., 2020; Kramer et al., 2020). Furthermore, the characteristics of the content to be observed apparently influence the acquisition of teachers' assessment skills (Hoppe, Renkl, & Rieß, 2020). To our knowledge, this assumption has not been systematically addressed with regard to the conceptual complexity of student thinking about science.

[1]University of Education Freiburg, Germany
[2]Technical University of Munich, Germany
[3]University of Freiburg, Germany

Corresponding Author:
Tobias Hoppe, University of Education Freiburg, Kunzenweg 21, 79117 Freiburg, Germany.
Email: tobias.hoppe@ph-freiburg.de

With the goal to develop teacher trainings that promote core teaching practices, McDonald et al. (2013) highlight the need for "a deep understanding of how people learn to enact ambitious professional practice" (p. 379) in practice-based learning environments. More research is needed on how specific characteristics of learning environments for teacher professional education have an impact on their learning practice (Kavanagh et al., 2020). In this study, we present a concept of how the assessment of student thinking about significant ecological concepts can be fostered in line with principles of practice-based learning (Grossman et al., 2009; McDonald et al., 2013). Within a teacher training session, we experimentally varied essential components of the learning environment to gain a deeper understanding of how instructional elements in practice-based learning and different characteristics of student thinking affect the acquisition of preservice science teachers' assessment skills.

Teacher Assessment Skills: A Prerequisite for Responsive Teaching

Students build their knowledge on pre-existing ideas. To provide all students with appropriate learning opportunities, teachers need to address student thinking in the course of a lesson (Shavelson et al., 2008). This approach of responsive or adaptive teaching is supported by a large body of evidence (Decristan et al., 2015; Kavanagh et al., 2020; Ruiz-Primo & Furtak, 2007). Given that adaptive teaching is responsive to individual students' learning requirements, it is seen as an essential means to achieve equity among learners (Windschitl & Barton, 2016). Following a Delphi expert panel approach, Kloser (2014) unsurprisingly identified the teaching practice of eliciting, assessing, and using student thinking as a core practice in teaching science. Kloser described this practice as teachers engaging in both formal and informal ways of probing student thinking and identifying students' conceptions and mental models of the material world and scientific practices using a variety of assessment practices. Science teachers should then use the information on their students' thinking to guide further instruction. These activities are considered to be common formative assessment practices during classroom interaction (Furtak, Kiemer, et al., 2016). Teacher judgments on student thinking form the basis of all further steps of formative assessment because teachers can only act on what they perceive, and the way teachers interpret student thinking determines how they may respond, give feedback to a student, or ask a question to further explore student thinking (Ruiz-Primo & Furtak, 2007; Schoenfeld, 2011; van Es & Sherin, 2021).

Situations in which teachers assess and use student thinking may be planned by the teacher, yet they often arise spontaneously and unplanned, for example, when a teacher overhears student statements during group work (Shavelson et al., 2008). In both cases, making judgments about student thinking requires teachers to engage immediately in assessment processes that involve the assessment skills of noticing and interpreting (Blömeke et al., 2015; Heitzmann et al., 2019). Noticing refers to teachers being able to draw their attention to relevant student statements among all other events happening during a school lesson. Interpreting refers to how teachers make sense of what they have noticed (Farrell et al., 2022). These cognitive skills can be considered integral to teachers' assessment competence (Blömeke et al., 2015; Heitzmann et al., 2019; Hoppe, Renkl, & Rieß, 2020; Loibl et al., 2020) and have also been described in research on teachers' professional vision (Blomberg et al., 2011) and teacher noticing (van Es & Sherin, 2002). The closely connected processes of noticing and interpreting lead to a more or less adequate judgment about student thinking (Loibl et al., 2020).

Essential indications for student thinking are provided by their verbal statements. To reach a hypothetical understanding of student thinking, the students' statements must be cautiously interpreted (Duit et al., 2008). Adequate interpretations of student thinking impose a particular challenge on teachers (van Es, 2011). The pedagogical task for science teachers subsequent to a judgment entails linking student thinking based on their everyday experience to scientific knowledge (Gebhard et al., 2017). Teachers may build bridges from (or draw contrasts between) intuitive student thinking to scientific knowledge, which enhances students' ability to reconstruct their conceptions toward scientifically valid perspectives (Duit et al., 2008). These strategies may be implemented, for example, by direct instruction or facilitating classroom discourse. Interpreting student thinking in a way that can inform such pedagogical decisions is difficult (Furtak, Kiemer, et al., 2016). Less experienced teachers tend to interpret student thinking according to the student's general behavior or aspects of classroom management, for example, when teachers remark that a student hardly participates in a discussion (Schoenfeld, 2011; van Es, 2011). However, even when referring to subject-specific aspects in student statements, teachers often tend to evaluate student thinking as only right or wrong (see Chi et al., 2004; Furtak, Kiemer, et al., 2016; Ruiz-Primo & Furtak, 2007). To guide further instruction and provide students with appropriate learning opportunities, teachers must gain an understanding of the specific conceptions that underlie students' statements and comprehend the way students think (Furtak, Kiemer, et al., 2016; Gropengießer & Marohn, 2018; van Es & Sherin, 2021).

Common Student Conceptions: Evidence-Based Representations of Student Thinking

Specific characteristics of student thinking as the object of assessment are likely to determine the quality of teachers' judgments (Hoppe, Renkl, & Rieß, 2020; Schrader &

Praetorius, 2018). Student thinking on scientific content varies greatly (Duit, 2009). To understand typical patterns of student thinking, researchers in science education have described student conceptions of scientific principles, phenomena, and concepts for decades. Student conceptions describe subjective mental constructions and have also been designated as, for example, student misconceptions, alternative conceptions, or naïve theories. When researchers aim for a description of student conceptions, they usually collect and evaluate qualitative data using, for example, group discussions. Although some student conceptions of varying scientific topics (e.g., anthropomorphisms) might occur, most student conceptions are domain-specific, that is, they differ on the scientific topic they refer to (e.g., blood circulation, or decomposition processes; Hammann & Asshoff, 2015).

Furthermore, student conceptions can differ in the complexity of their conceptual construction (Gropengießer, 2006). Gropengießer distinguishes, for example, between concepts, constructs, and principles. A *concept* represents the simplest form of an idea and refers to things, processes, or general referents (Chi et al., 1994). They are expressed by words or terms (Gropengießer & Marohn, 2018). When learners relate a term to one or other terms, a *construct* is built in the learner's mind. Constructs refer to concrete facts and can be expressed by a single assertion or sentence (Gropengießer & Marohn, 2018), for example, blood is primarily contained in blood vessels (Chi et al., 1994). *Principles* are particularly complex ideas in which different *constructs* are related to each other. Based on a large body of research, teachers and teacher educators might refer to comprehensive collections of common student conceptions about the interpretation of student thinking as a basis for further instruction (Duit, 2009). In such collections, common conceptions are usually described giving a brief description of the conception as well as exemplary student statements indicating a certain conception (Pine et al., 2001). In various prominent models of pedagogical content knowledge (PCK), knowledge about student conceptions is considered to be a key component of teachers' professional knowledge (Loewenberg Ball et al., 2008; Shulman, 1984).

Understanding student conceptions requires complex reconstructive analysis processes because student statements must be interpreted by inference. This processing is difficult in the middle of a school lesson when teachers have no prior knowledge on potential student conceptions (Dannemann, 2020). Fortunately, many typical student conceptions have been described in empirical research for a variety of scientific topics. Descriptions of student conceptions usually contain an applicable concept name, a definition, and anchor examples (e.g., Hammann & Asshoff, 2015). These descriptions of typical student conceptions can serve as categories for teachers' on-the-fly judgments of student thinking (Taber, 2017). However, an evidence-based category is, in most cases, an abbreviation of what students actually think, and does not cover all variants and deviations from this category (Dannemann, 2020). Therefore, a spontaneous judgment of student thinking based on such categories can only be considered a tentative but useful hypothesis, given the challenges of "on-the-fly" assessment of student conceptions (Heitzmann et al., 2019). Gaining a hypothetical understanding of how students think, teachers can then move on to adaptive instruction, or they can ask questions to validate their understanding of how they perceive a student's thinking (van Es & Sherin, 2021).

Practice-Based Learning: An Approach to Foster the Application of Practice-Oriented Knowledge

To educate prospective teachers successfully in acquiring practice-oriented knowledge, the framework for teaching practice developed by Grossman et al. (2009) suggests identifying common core practices of the teaching profession and decomposing them into individual components that can then be rehearsed in complexity-reduced settings that display only certain aspects of classroom practice. Common core practices of the teaching profession represent "broadly applicable instructional strategies known to foster important kinds of student engagement and learning" (Windschitl et al., 2012). Such practices occur frequently in teaching, and they preserve the complexity of teaching in the sense that learning opportunities must be tailored to the specific needs of particular students (Grossman et al., 2009; Windschitl & Barton, 2016). For science teaching, a Delphi study highlighted eliciting, assessing, and using student thinking about science as a particularly significant practice (Kloser, 2014). Noticing and interpreting student conceptions which underlie student statements represent an essential component within this core teaching practice (Chi et al., 2004; Furtak, Thompson & van Es, 2016; Ruiz-Primo, 2011; Schoenfeld, 2011).

Teaching preservice teachers components of a more complex practice such as noticing and interpreting student thinking does not ensure they can adequately use these practices when they enact complete lessons (Kennedy, 2016). In practice-based teacher education, it is essential that teachers learn to connect those practices to specific purposes which relate to students, context, and content (Hauser & Schneider Kavanagh, 2019). McDonald et al. (2013) suggested organizing practice-based learning in a cycle in which preservice teachers can rehearse components of core practices engaging in activities with increasing complexity. The authors stress the need for the development and identification of pedagogies (e.g., video analysis), which teacher educators enact to support teachers in learning practice. To design learning environments based on principles of practice-based learning that foster specific skills, an important step is to assess how different components of a learning setting contribute to the acquisition of teacher skills. The question of which representations of practice can be used and implemented in teacher training is particularly prevalent within practice-based

learning (Codreanu et al., 2020; Danielson et al., 2018; Grossman et al., 2009).

Written and Video Cases: Representations of Practice. Video and written cases are commonly used as representations of practices (Danielson et al., 2018). Both media types provide the opportunity for teacher students to engage in certain aspects of teaching practice without facing the pressure of a real classroom situation. Novice teachers are offered the chance to rehearse, revise, and retry certain aspects of practice (Grossman et al., 2009). Video cases are considered to be particularly authentic because they display many facets of practice. Yet, this complexity may lead to mental load on the viewers and be perceived as overtraining (Farrell et al., 2022). However, results in research on mental load caused by video viewing have not been consistent (Kramer et al., 2020). Written cases are considered to be less authentic because they display fewer facets of classroom practice (Danielson et al., 2018). The lower authenticity of written cases offers its own advantages (Grossman et al., 2009). For example, written cases allow participants to sequentially process the information presented (Gold et al., 2016).

Studies have shown that text and video cases have similar potential in promoting teacher skills in noticing and interpreting classroom situations (e.g., Hoppe, Renkl, & Rieß, 2020; Kramer et al., 2020; Schneider et al., 2016). Given that video and written cases offer different potentials and limitations, the question arises as to how both media types contribute to the acquisition of cognitive skills when used in combination. Learners might thus benefit from the complexity and richness of a classroom situation represented in videos as well as the opportunity for sequential processing of information that written cases offer. Research on the effects of media-combined learning environments for teachers has been very scarce. Kramer et al. (2020) found that compared to using transcripts only, a combination of videos and transcripts in fostering teachers' situation-specific skills of classroom management was particularly effective. Participants in the media-combined condition worked with transcripts first and then moved on to working with videos. The transcript and video sequence was not varied.

When evaluating the benefit of written and video cases in teacher training, motivational constructs are often measured in addition to the outcomes of the intended cognitive teacher skills (Codreanu et al., 2020). Given that videos represent many facets of real classroom practice, and they offer viewers the opportunity to become immersed in the classroom situation (Farrell et al., 2022). This motivational construct has been framed as *immersion* or *cognitive involvement* in video-based learning (Schneider et al., 2016; Seidel et al., 2011; Syring et al., 2015). For example, Syring et al. (2015) measured the extent that participants were engaged and involved in video and text-based case work on classroom management. They found that immersion, understood as the level of engagement and involvement in case work, increased slightly over time in the video group as well as the text group. However, immersion always remained higher in the video than in the text group. Participants seemed to be more involved in video cases than in text cases, which also led to slightly higher engagement in case work. These findings raise the question of whether, for example, viewing a video first may facilitate participants' immersion in a text case. When combining text and video cases in practice-based learning, an evaluation of participants' perceived immersion appears to be particularly fruitful, yet to our knowledge, it has not been investigated to date.

Rehearsals of Components of Practice. Practice-based learning suggests rehearsing a certain component of teacher practice and then moving on to the next step characterized by more features of authentic practice (Grossman et al., 2009; McDonald et al., 2013). Such an increase in authenticity, which Grossman et al. (2009) called approximations of practice, might involve a more complete or integrated representation of practice or fuller participation by a novice (Grossman et al., 2009). For example, participants might examine video exemplars and in a subsequent step rehearse what they observed in a live role-play. This procedure in practice-based learning implies the question of how much rehearsal is needed until participants are ready to move to the next step. When practicing assessment of student thinking, the teachers' learning progress likely depends on the scientific target content of student thinking as well as on the conceptual complexity of students' statements that are being noticed. Focusing on student conceptions about evolution, Fischer et al. (2021) found that trainee teachers assessed anthropomorphic misconceptions significantly more often than teleological misconceptions. Effects of the conceptual complexity of student thinking on teachers' assessment competences to the best of our knowledge have not been systematically tested and reported. Nonetheless, when developing pedagogies of how teachers can engage in teacher practices in successive steps, such as assessing student thinking, an evaluation of how much practice is needed with respect to varying content complexity is crucial. In a small sample study, Hoppe, Renkl, & Rieß (2020) found that learning to assess more complex student conceptions appears particularly challenging compared to assessing less complex student conceptions.

Aim of Study

The present study aimed to expand knowledge on the design of practice-based teacher trainings for fostering teachers' assessment skills of student thinking. We assessed how the use of different media types as representations of practice, the number of rehearsals, and different characteristics of student thinking that account for its complexity contribute to preservice teachers' acquisition of assessment skills. Furthermore, we examined how a combination of these

aspects in practice-based teacher trainings can advance teachers' on-the-fly judgments of student thinking. We addressed the following research questions:

Research Question 1: To what extent do different case types as representations of practice (video or written cases) and student conceptions of different complexity impact (a) the acquisition of preservice teachers' assessment skills and (b) preservice teachers' perceived immersion in a first rehearsal?

Our first research question aimed to validate previous findings. Correspondingly, we assumed that video and written cases are both suitable to promote teachers' acquisition of assessment skills. With respect to the complexity of student conceptions, we expected that the acquisition of assessment skills would occur differently because student conceptions of different conceptual complexity represent different requirements for teachers' noticing and interpreting. Moreover, video cases are often assumed to bear particular potential because they provide an opportunity for participants to become immersed in the situation, but previous research shows no clear advantage of videos for participants' perceived immersion in a case.

Research Question 2: To what extent does a media-combined sequential variation of video and written cases and student conceptions of different complexity impact (a) the acquisition of preservice teachers' assessment skills, and (b) preservice teachers' perceived immersion in a second rehearsal?

This research question was explorative given its novelty. With regard to the teaching principle of increasing complexity, preservice teachers might benefit from practicing the assessment of less complex content using less complex representations of practice (written cases) before moving on to more complex content using a more complex representation of practice (video cases). With regard to participants' perceived immersion it is novel and of particular interest to examine how they perceive their immersion in a media-combined training.

Research Question 3: Does learning to assess complex student conceptions require more rehearsals than learning to assess less complex student conceptions?

Student conceptions vary not only on the scientific topic they refer to but also on the complexity of their conceptual construction (Chi et al., 1994; Gropengießer & Marohn, 2018). The question is whether the number of necessary rehearsals depends on different degrees of conceptual complexity in student conceptions. This research question also has not been addressed in the context of on-the-fly assessment. Expecting that assessing more complex conceptions would be more difficult, we assumed that participants benefit from more rehearsals.

Method

Participants and Design

Participants were recruited within a course in a Biology education master's program for preservice teachers at two universities. The sample consisted of 104 participants (81 female, 23 male). The average age was 24.75 ($SD = 3.48$). Sixteen preservice teachers were enrolled in a program for elementary school teachers, and 88 were enrolled in a program for secondary school teachers. All participants had completed a bachelor's program in Biology covering the ecological subject area under investigation in this study. Furthermore, all participants had taken part in a mandatory teaching internship integrated into the preceding bachelor's program. The structure of the study program for preservice Biology education teachers is largely similar to study programs at universities throughout Germany. Participants were not familiar with specific ecological student conceptions.

The intervention study followed an experimental pre-posttest design including a further interim measurement time. The individual preservice teachers were randomized to experimental conditions (see Table 1). All participants in experimental groups took part in a practice-based assessment training comprised of two rehearsal sessions. Within the experimental conditions, we varied media type (video or text) as well as content (student conceptions of different complexity). After an interim measurement time, media type and content were switched for each group in the second rehearsal (e.g., video case with less complex content during the first rehearsal, then written case with complex content during second rehearsal). The student conceptions represented in the classroom cases referred to basic ecological concepts from the overarching topical domain of material cycles in ecosystems. A control condition had access to the same information about specific ecological student conceptions as the experimental conditions, but participants did not rehearse assessing student conceptions. Instead, they took part in lesson planning activities addressing those specific student conceptions. More precisely, participants in the control condition were asked to think about learning activities that build on those specific student conceptions and allow students to move their thinking forward (e.g., generating cognitive conflict in the mind of a student).

Development of Video and Written Cases to Promote Assessment Skills

To foster preservice teachers' assessment skills, the research team developed classroom cases following manuals that depict the production of scripted yet authentic classroom vignettes (Piwowar et al., 2017; Seidel et al.,

Table 1. Experimental Groups and Control Group (N = 104).

Group (IV)	*n*	Media type and content during the first rehearsal	Media type and content during the second rehearsal
VideoC/Text-C	19	Videos, complex student conceptions (C)	Transcripts, less complex student conceptions (-C)
TextC/Video-C	20	Transcripts, complex student conceptions (C)	Videos, less complex student conceptions
Video-C/TextC	20	Videos, less complex student conceptions (-C)	Transcripts, complex student conceptions (C)
Text-C/VideoC	20	Transcripts, less complex student conceptions (-C)	Videos, complex student conceptions (C)
Control group	25	C/-C (no videos or transcripts)	C/-C (no videos or transcripts)

Note. Complex student conceptions (C) can be classified as principles (Gropengießer & Marohn, 2018) and topically refer to the importance of plants in ecosystems. Less complex student conceptions (-C) can be classified as constructs and topically refer to the decomposition of matter.

Lina:	(reads task out loud) Where does the change in leaves come from? Give reasons for your statements.
Adrian:	I guess it's the frost and the wind.
Lina:	Yes, I just think that small animals, for example, ants cut up the leaves and then take them to their burrows and that is how the leaves get those holes and so on.
Danielle:	(looks at Lina) Isn't this task similar to the last one? (Lina raises her shoulders) You're saying exactly the same thing.
Jamal:	Well, I definitely think that nature just made it that way and that it's nature's intention that leaves just change like that in the course of a year.

Figure 1. Transcribed Episode From a Video Vignette.
Note. Students discuss the causes of decomposition processes.

2022). The video cases were approximately 2 min long. For each topic domain (importance of plants in ecosystems and decomposition processes), three vignettes had been produced. In each vignette, a short group work episode (science classroom) was presented. Four students worked on their respective topic, that is, importance of plants in ecosystems or decomposition processes. Students' verbal statements revealed underlying student conceptions in these domains. The student conceptions had been selected to represent common student conceptions, as identified in various empirical studies and among students at primary and lower secondary levels (Hammann & Asshoff, 2015). The selected student conceptions on the importance of plants in ecosystems can be classified as complex *principles* (C), whereas the student conceptions on decomposition processes can be classified as less complex *concepts* (-C) (Gropengießer & Marohn, 2018). Thus, the content of the classroom vignettes differed with respect to the complexity of student conceptions and the topical focus.

To develop comparable text cases, the video cases were transcribed. Complementary non-verbal information (e.g., students' gestures) had been added. Thus, we developed six cases in text format, each presenting about 2,000 text signs. Video and text cases contained additional information on school type, class level, and topic of the lesson.

Measures and Instruments

Assessment Skills. To assess the development of preservice teachers' assessment skills, participants took part in a video-based test prior to, within, and after the intervention. The same instrument was used at each measurement point. The video-based test consisted of two scripted videos (each approx. 2 min) displaying a science classroom scene at a lower secondary level (see Figure 1). Similar to the cases used during the intervention, a group of four students worked on tasks related to the topic domain of material cycles in ecosystems. In a group discussion, they expressed their conceptions about the subject matter. The test consisted of two parts. Each of the two videos was linked to one of the student conception complexity levels and to a specific topic domain. The first part of the video test (Part A) represented complex student conceptions (*principles*) about the importance of plants in ecosystems. The second part of the video test (Part B) represented less complex student conceptions (*concepts*) about decomposition processes. While less complex student conceptions on decomposition processes can be concluded from a single statement, for assessing the more complex student conceptions about the importance of plants in ecosystems, different statements must be related to each other. The videos represented a total of 23 situations in which student conceptions were noticeable.

Table 2. Description of the Scale for Immersion: Number of Items (N) and Sample Items.

Scale	*N*	Items
Immersion	7	When I was working on the cases I was fully involved. When I was working on the cases I wanted to know more. When I was working on the cases I found the sections boring.

Note. Likert-type response scale from 1 (disagree) to 4 (agree).

In addition to the video, participants received information about the classroom sequence (e.g., student names, class level, and learning materials) to provide a more comprehensive picture of the assessment situation and to allocate observations to certain students. The classroom cases used for the test differed from the cases used in the intervention in the group of students and the tasks on which the students worked. Participants were asked to analyze the videos with a focus on what seemed relevant for further teaching and learning. We used an open-question format which had already been used for determining various competence levels in previous studies (e.g., Kersting, 2008). To represent the fleetingness of the actual classroom situation, the videos could only be paused to note down observations. Single episodes or complete videos could not be watched again.

The quality of teachers' assessment skills was operationalized as the level of noticing and interpreting student statements about underlying conceptions. We coded participants' responses following the qualitative content analysis (Mayring, 2015). Categories had been formed deductively and inductively. First, we coded preservice teachers' noticing in the sense that participants selected relevant student expressions that hint at the underlying conceptions. Selecting means that participants had notes about observations with reference to a specific anchor in the video or written case. We further determined the quality of participants' observations in terms of their interpretations of presumable underlying student conceptions. For example, we distinguished whether student expressions were evaluated by their scientific correctness or whether an attempt had been made to understand the students' way of thinking (Furtak, Kiemer, et al., 2016; van Es, 2011). Thus, a total of 18 categories emerged to describe the participants' responses. These 18 categories were transferred into a six-level quasi-interval scale which was validated by three experts who assigned the categories to the values 1 to 6 (1 = very low level of assessment skills, 6 = very high level of assessment skills; intraclass correlation coefficient [ICC] = .938). Lower levels describe, for example, merely descriptive comments or comments that interpret student expressions about general student behavior. Lower levels of assessment skills further describe comments that evaluate student expressions as right or wrong but lack attempts to understand the students' ways of thinking. Higher levels of assessment skills describe attempts to gain an understanding of the specific conceptions that underlie students' statements and to comprehend the way students think. Each participant's response that referred to a relevant passage in the video was assigned a value from 1 to 6. The scores were summed for each part of the video test. This procedure resulted in a sum score representing a comprehensive measure of assessment skills based on the preservice teachers' proficiency. This measure includes both cognitive processes of noticing and interpreting at different degrees of adequacy. To determine the reliability of measuring assessment skills using the vignette test, Cronbach's alpha was calculated. The reliability was acceptable for Part A of the video test (Cronbach's alpha = .72) and excellent for Part B (Cronbach's alpha = .91).

Immersion. To assess the preservice teachers' perceived immersion while rehearsing assessment of student conceptions using video and written cases, participants rated their perception on a scale for immersion immediately after each rehearsal session (see Table 2). The scale consisted of seven items and was adapted from Syring et al. (2015). The internal consistency of the questionnaire was satisfying, with Cronbach's alpha for perceived immersion = .75.

Procedure. All participants attended two lecture-style sessions each lasting for 90 min (see Table 3). In the first session, ecological content knowledge relevant to assessment (e.g., decomposition processes, plant ecology) was addressed. The second session addressed general PCK on student conceptions (e.g., the role of student conceptions for student learning). The participants were further introduced to specific PCK on typical student conceptions in the content domains of ecology. This introduction to typical ecological student conceptions took place within two consecutive 90-min rehearsal sessions.

In an example-based learning environment, participants were able to analyze classroom cases (video or written format) on the student conceptions that were expressed. Initially, participants were introduced to solution examples as to how student expressions can be successfully assessed according to the underlying student conceptions. At the same time, participants received specific PCK about common student conceptions of either complex conceptions about the importance of plants in ecosystems or less complex conceptions about decomposition processes (see Table 4). For the first video/written case, participants were given examples of what ecological student conceptions can be observed in the classroom sequence. They were then asked to watch the video again and comprehend the solution examples. Participants in the text group were asked to read the text again. For the next two videos/written cases, participants were asked to assess student conceptions themselves. For each video/written case,

Table 3. Procedure of Tests, Lectures, and Practice-Based Trainings.

Group	*n*	Variation of media type and content complexity in practice-based trainings				
VideoC/Text-C	19	Pretest Lectures	VideoC	Interim test	Text-C	Posttest
TextC/Video-C	20		TextC		Video-C	
Video-C/TextC	20		Video-C		TextC	
Text-C/VideoC	20		Text-C		VideoC	
Control group	25		C/-C		C/-C	

Note. Pedagogical content knowledge (PCK) about common ecological student conceptions was taught within the case-based trainings. Complex student conceptions (C) referred to the topic domain of the importance of plants in ecosystems. Less complex student conceptions (-C) referred to the topic domain of the decomposition of matter. The control group (CG) received the same knowledge about common ecological student conceptions in both content areas. Prior to this, all participants were taught content knowledge and general PCK about student conception in two lecture-style sessions.

Table 4. Student Conceptions.

Category: student conceptions about the causes of decomposition of plant and animal matter		
Concept	Definition	Anchor examples
Decomposition due to the influence of abiotic factors (Hammann & Asshoff, 2015)	Decomposition processes are primarily attributed to the influence of abiotic factors (e.g., air, sun, cold, moisture).	"Hmm. . . I think frost decomposes the leaves." ". . . more like by the weather."
Decomposition by the action of macroscopically visible organisms (Hammann & Asshoff, 2015)	The cause of decomposition processes is seen exclusively in the action of macroscopically visible organisms (e.g., cows, insects).	"Snails and beetles maybe. . . they gnaw on the fruit. The sheep also do that." ". . . because little bugs then nibble on it. And worms."

Note. Examples of the information about common student conceptions were given to both the experimental conditions as well as the control condition.

participants were able to compare their judgments to an expert solution. Thus, participants were asked to practice applying their knowledge about specific student conceptions using a total of three classroom cases during the first rehearsal. After the interim tests, three further classroom cases followed during the second rehearsal. These practice-based teacher trainings were designed as prestructured group work and could be completed independently by the participants. The purpose of this design was to minimize the influence of the lecturers on the results of the different training conditions. Other conditions (e.g., learning time and formulation of the tasks) were kept constant across the groups.

In these rehearsals using representations of practice, we varied (a) the conceptual complexity, (b) the media type (video or written case), and (c) the sequence in which case type and content were presented within the two rehearsal sessions. Thus, for the second rehearsal phase, we exchanged case types and student conceptions in their complexity. For example, if participants had used video cases in the first phase, they received written cases in the second phase. If participants had rehearsed assessing complex student conceptions in the first phase, they rehearsed assessing less complex student conceptions in the second phase. The formal structure of the two rehearsal phases did not differ. After each rehearsal cycle, all participants in the experimental conditions were asked to report perceived immersion during enactment.

The sessions of a control condition were held chronologically parallel to the experimental conditions. Participants in the control condition received the same information about typical ecological student conceptions, including anchor examples as the experimental conditions (see Figure 2). Participants in the control condition included no rehearsal applying this knowledge using written and video cases, but they were asked to plan lessons that address ecological student conceptions. Thus, participants had an opportunity to process knowledge about typical student conceptions in-depth, but they were not able to enact such knowledge in an on-the-fly assessment activity using representations of practice.

Results

Table 5 displays the means and standard deviations of key variables in all experimental conditions. An alpha level of .05 was used for all statistical tests. We used partial η^2 as an effect size measure—qualifying values <0.06 as small effects, values in the range between 0.06 and 0.13 as medium effects, and values >0.13 as large effects.

Pre-analysis

We calculated analysis of variances (ANOVAs) to determine whether there were any statistically significant differences between groups. We found no difference between groups in

Table 5. Means and SD of Assessment Skills for Complex Conceptions (C) and Less Complex Conceptions (-C) at Three Measurement Times.

Variable	Experimental condition	*n*	*M*	*SD*
Assessment skills Part A: C (Pretest)	1 Video/C	19	3.37	2.73
	2 Text/C	20	5.20	4.25
	3 Video/-C	20	5.55	4.62
	4 Text/-C	20	5.50	6.26
	CG C/-C	25	4.76	3.79
Assessment skills Part B: -C (Pretest)	1 Video/C	19	3.84	3.85
	2 Text/C	20	7.40	7.44
	3 Video/-C	20	7.60	8.27
	4 Text/-C	20	9.65	9.26
	CG C/-C	25	5.40	5.13
Assessment skills Part A: C (Interim test)	1 Video/C	19	29.16	10.91
	2 Text/C	20	27.20	9.31
	3 Video/-C	20	21.75	11.65
	4 Text/-C	20	17.65	9.75
	CG C/-C	25	20.44	9.95
Assessment skills Part B: -C (Interim test)	1 Video/C	19	31.47	8.08
	2 Text/C	20	32.75	7.82
	3 Video/-C	20	35.50	7.73
	4 Text/-C	20	33.75	8.83
	CG C/-C	25	22.48	11.80
Assessment skills Part A: C (Posttest)	1 Video/C	19	18.21	13.31
	2 Text/C	20	16.40	13.22
	3 Video/-C	20	44.65	12.24
	4 Text/-C	20	45.80	10.58
	CG C/-C	25	22.56	18.01
Assessment skills Part B: -C (Posttest)	1 Video/C	19	47.68	8.39
	2 Text/C	20	49.00	13.58
	3 Video/-C	20	50.20	8.07
	4 Text/-C	20	51.90	7.97
	CG C/-C	25	29.08	19.23

Note. Part A of the video test displayed student statements indicating complex student conceptions (C) about the importance of plants in ecosystems. Part B of the video test displayed student statements indicating less complex student conceptions (-C) about the decomposition processes.

assessment skills or demographic variables such as age, enrollment in study programs, grades in biology, number of semesters, and teaching experiences (all $ps > 0.20$).

Research Question 1: To What Extent Do Different Case Types as Representations of Practice (Video or Written Cases) and Student Conceptions of Different Complexity Impact (a) the Acquisition of Preservice Teachers' Assessment Skills and (b) Preservice Teachers' Perceived Immersion in a First Rehearsal?

To examine the effectiveness of case-based training using different case formats and content of different complexity, we compared preservice teachers' assessment skills between each experimental group and between experimental groups and a control condition. Results were generated separately for the assessment of complex content (A) and the assessment of less complex content (B) (see Table 5). After the first rehearsal, a rmANOVA showed significant differences in preservice teachers' assessment skills for complex student conceptions (Part A) in comparison to pretest results, $F(4, 99) = 4.77$, $p < .001$, highlighting an effect size of $\eta^2 = 0.16$. Bonferroni-adjusted post hoc tests revealed that none of the experimental groups outperformed participants in the control condition after practicing assessment of complex student conceptions. Analyses further revealed statistically significant differences in assessment skills between groups that had rehearsed assessing complex student conceptions and the Text-C/VideoC group, which had rehearsed assessing less complex student conceptions (with a different topical focus), $p = .007$ (Video/C) and $p = .034$ (Text/C).

We further compared preservice teachers' achievement in assessing less complex (Part B) after the first rehearsal. Taking into account the results for pretest and interim test, a rmANOVA determined statistically significant differences between groups for the mean quality of assessment skills between groups, $F(4, 99) = 18.61$, $p < .001$, with an effect size of $\eta^2 = 0.43$. Post hoc analyses showed that groups that had rehearsed assessing less complex student conceptions outperformed the control condition (Bonferroni-adjusted $ps < .001$) and groups that had rehearsed assessing complex student conceptions (Bonferroni-adjusted $ps < .001$). We found no evidence indicating a higher benefit from either the video or the text condition. Comparisons with the control condition revealed that groups benefited significantly from practice-based training only if they had rehearsed assessing less complex conceptions and only for assessing less complex conceptions with the same topical focus (Part B).

On average, participants in all experimental groups evaluated their immersion in the upper half of the Likert-type response scale during the first rehearsal (see Table 6). An ANOVA revealed no significant differences between groups, $F(3, 70) = 1.16$, $p = .330$.

Research Question 2: To What Extent Does a Media-Combined Sequential Variation of Video and Written Cases and Student Conceptions of Different Complexity Impact (a) the Acquisition of Preservice Teachers' Assessment Skills and (b) Preservice Teachers' Perceived Immersion in a Second Rehearsal?

Case format and content were changed in the second rehearsal. For assessing complex student conceptions (Part A), an rANOVA showed significant differences in preservice teachers'

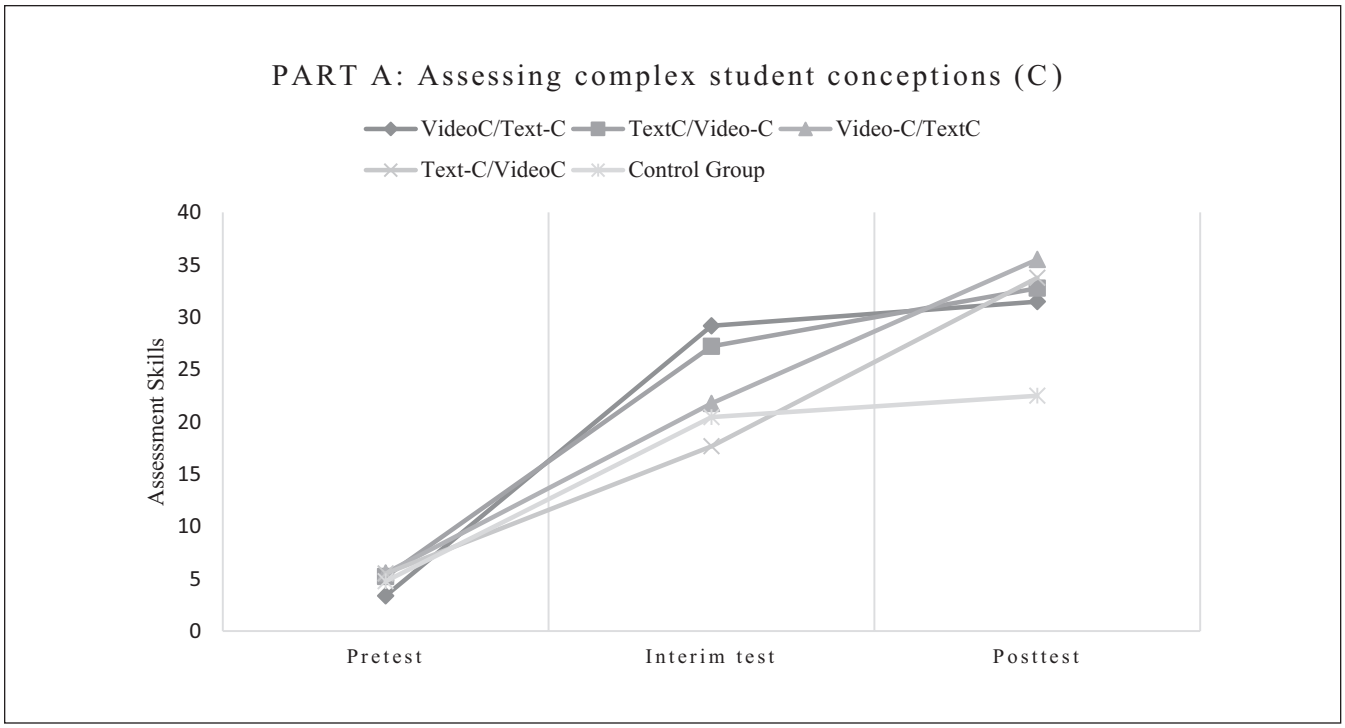

Figure 2. Results for Assessing Complex Student Conceptions.
Note. The Video and Text designations indicate whether the group had practiced with classroom videos or transcripts. C and -C indicate whether the group had practiced assessing complex student conceptions (C) focusing on the importance of plants in ecosystems or less complex student conceptions (-C) focusing on the decomposition of matter. For example, the group VideoC/Text-C had practiced assessing complex student conceptions using classroom videos during the first rehearsal, and less complex student conceptions using classroom transcripts during the second rehearsal.

assessment skills between the experimental conditions and the control condition, $F(4, 99) = 6.23$, $p < .001$), highlighting an effect size of $\eta^2 = 0.21$ (see Figure 2). All experimental conditions outperformed the control condition, $p = .014$ (VideoC/Text-C), $p = .003$ (TextC/Video-C), $p < .001$ (Video-C/TextC), $p = .001$ (Text-C/VideoC). Participants' performance was not statistically different between experimental conditions. Analyses of the preservice teachers' assessment skills for less complex student conceptions revealed a similar picture. A rmANOVA showed significant differences, $F(4, 99) = 10.01$, $p < .001$, with an effect size of $\eta^2 = 0.29$ (see Figure 3). Bonferroni-adjusted post hoc tests revealed that all experimental conditions outperformed the control condition, all ps $< .001$, and no statistically significant differences between experimental groups were found.

After the second rehearsal phase, participants evaluated their immersion again in the upper half of the Likert-type response scale (see Table 6). Yet, an ANOVA revealed significant differences between participants, $F(3, 65) = 4.08$, Bonferroni-adjusted $p = .010$, highlighting an effect size of $\eta^2 = 0.16$. Participants in the VideoC/Text-C group reported significantly lower scores for their immersion than participants in the Video-C/TextC group. t-Tests further showed that compared to the first phase of rehearsals, participants in the VideoC/Text-C reported significantly lower scores for the second rehearsal phase, $t(15) = 4.06$, $p < .001$. For all other experimental groups, no statistically significant differences were found between their evaluation or the first and the second rehearsal phase.

Research Question 3: Does Learning to Assess Complex Student Conceptions Require More Rehearsals Than Learning to Assess Less Complex Student Conceptions?

To address this research question, we refer to the analyses presented in the previous sections Research Questions 1 and 2. We compared the experimental groups' performances to the performance of a control condition (see Table 5). Participants in all experimental groups outperformed participants in the control condition in assessing complex student conceptions only after a second rehearsal. Participants in the experimental conditions outperformed participants in the control group in assessing less complex student conceptions after the first rehearsal. However, participants in the experimental conditions only showed better assessment skills than the control condition, if they had rehearsed assessing less complex student conceptions (which had a different topical focus than complex student conceptions).

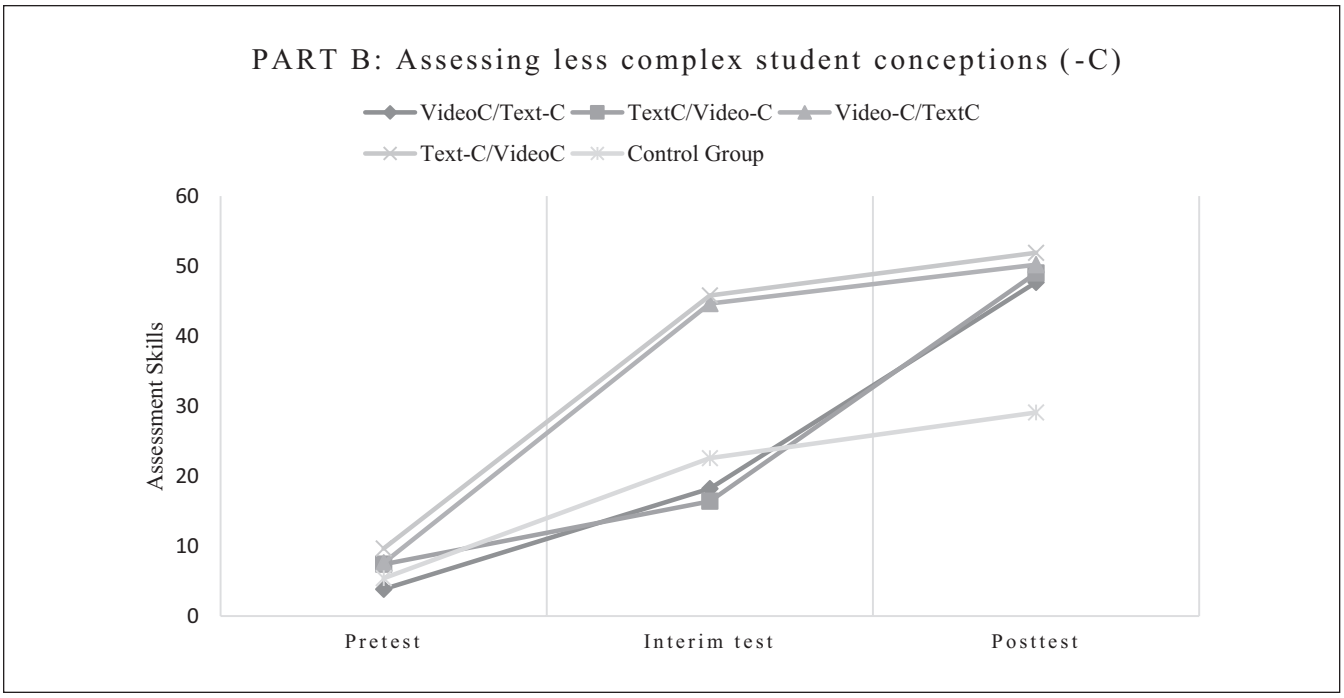

Figure 3. Results for Assessing Less Complex Student Conceptions

Note. The Video and Text designations indicate whether the group had practiced with classroom videos or transcripts. C and -C indicate whether the group had practiced assessing complex student conceptions (C) focusing on the importance of plants in ecosystems, or less complex student conceptions (-C) focusing on the decomposition of matter. For example, the group VideoC/Text-C had practiced assessing complex student conceptions using classroom videos during the first rehearsal, and less complex student conceptions using classroom transcripts during the second rehearsal.

Table 6. *Means (SD) of Immersion After Both Rehearsal Sessions.*

Construct	After first rehearsal					After second rehearsal				
	1 Video/C	2 Text/C	3 Video/-C	4 Text/-C	CG C/-C	1 Text/-C	2 Video/-C	3 Text/C	4 Video/C	CG C/-C
Immersion	3.32 (0.39)	3.32 (0.48)	3.46 (0.42)	3.21 (0.33)	–	2.91* (0.32)	3.16 (0.52)	3.39* (0.34)	3.24 (0.45)	–

Note. *Significance of cross-sectional mean difference. Significant at an alpha level of $p < .01$. Likert-type response scale from 1 (disagree) to 4 (agree).

Discussion

This article focuses on how preservice science teachers can be supported in developing cognitive skills for assessing student thinking when interacting with students. Making such on-the-fly judgments of student conceptions informs teachers' pedagogical decisions on further instruction in response to their students' thinking. We focused on teachers' on-the-fly assessment processes as a key decomposition within formative assessment activities. Following the teaching practice approach, we investigated the effects of media type, content complexity, and a sequential variation of those components on the development of preservice teachers' assessment skills.

In Research Question 1, preservice teachers' acquisition of assessment skills and perceived immersion within a video or text-based training was addressed. We found no evidence that either video or written cases are more beneficial in a single rehearsal. This finding is in line with previous research (Friesen & Kuntze, 2018). When choosing and developing suitable representations of practice, educators need to consider which facets of practice are made visible in the representation (Danielson et al., 2018). When assessing student thinking, noticing and interpreting what students say is important. Unlike non-verbal features such as mimics and gestures, students' verbal statements might be well represented in both video and written cases. Although after a single rehearsal, both case types appear to be suitable for increasing participants' on-the-fly assessment skills, the conceptual complexity of the content has a notable impact on the acquisition of these skills. Only experimental groups that had practiced assessing less complex student conceptions differed significantly from the control group's performance. Conclusively, differences in the complexity of student thinking require an adaption of the learning environment.

Apart from the evaluation of preservice teachers' assessment skills, we evaluated participants' perceived immersion during rehearsals as a motivational construct. Participants in all conditions reported high values for perceived immersion while using written or video cases to practice assessing student thinking. Thus, we found no advantage for participants' immersion with any particular media type and content after one rehearsal. Participants were able to instead achieve a high level of cognitive involvement in video and written cases.

Research Question 2 addressed the impact of a sequential variation of video and written cases and content of different conceptual complexity on the development of assessment skills and participants' immersion during rehearsals. Analyses showed that all experimental groups substantially benefited in the development of assessment skills from consecutive rehearsals in all variations. Unlike after the first rehearsal, all experimental groups after the second rehearsal differed from the control condition, which had received the same PCK about common student conceptions about ecology but had no opportunity to apply this knowledge in text-based and video-based teacher training. This finding underlines the benefit of practice-based learning for teachers' professional development in the case of on-the-fly judgments of student thinking.

Our findings indicate that within the approach of teaching practice using text and video cases to develop teachers' assessment skills, different paths may prove to be effective. To confirm this assumption, further research is needed on the variation of media types and different content qualities in learning sequences for practice-based teacher education. These findings for the practicalities of science teacher education are not trivial. Results may encourage teacher educators to implement an element of choice when preservice teachers learn to assess student thinking in practice-based learning environments. Participants may choose a case format and content they prefer for rehearsals. By allowing prospective teachers a high degree of control over the learning process, their motivation to acquire assessment skills will likely be sustained, which might lead to enhanced learning outcomes (Scheiter, 2014).

Apart from analyzing the development of assessment skills, an evaluation of how participants reported their immersion during rehearsals contributes to a more nuanced understanding of the impact of using different media types that represent content of different complexity in learning sequences during teacher trainings. For their perceived immersion, participants in all experimental conditions reported high values for both rehearsals. However, results for the immersion of the experimental group that assessed complex student conceptions using videos as the more complex case format in the first rehearsal and then moved to assess less complex student conceptions using written cases as the less complex case format in the second rehearsal, offer a more nuanced evaluation of the intervention. This result partially differs from the findings of Syring et al. (2015). They compared participants' immersion in video and written cases at three measurement times and found a significant difference in participants' immersion in favor of the video condition at the first measurement. In the other two measurements, the differences were not significant. We suggest that not only the media type might influence immersion but characteristics of the content in interaction with the media type could also play a role. According to our theoretical considerations, the experimental condition that reported significantly lower values for perceived immersion was the only group that experienced no progression in complexity of either content or case type. This result suggests that for participants' motivation, an element of progression in the complexity of the media type or content complexity when creating a series of rehearsals could be an important factor to consider. To test this hypothesis and to further explore how media type and choice of content affect participants' motivation further research is needed.

Generally, teacher trainings designed for practicing evidence-based assessment of student thinking based on categories derived from common student conceptions helps teachers develop first reasonable hypotheses about their students' thinking as a preliminary condition for further instruction. Apart from serving as an anchor for adaptive teaching or as further questions to explore student thinking, category-based interpretations of student conceptions contribute to the additional goal of developing a common language in the professionalization of teaching (Grossman et al., 2009; Heitzmann et al., 2019; Windschitl & Barton, 2016). In the context of formative assessment, being able to name specific *diagnoses* of student thinking might enhance the development and the application of *treatments* as teachers' instructional decisions tailored to specific forms of student thinking (Taber, 2017).

Although different paths appear to provide an opportunity for successful learning in multiple rehearsals, the question remains as to how much practice is needed to learn to assess a specific target content. In Research Question 3, the number of rehearsals needed for practicing the assessment of student conceptions that displayed different degrees of conceptual complexity was addressed. Comparisons with the control condition showed that for assessing less complex student conceptions, participants highly benefited from a single rehearsal session. For assessing complex student conceptions, the experimental groups only differed from the control group's performance after the second rehearsal. This finding suggests that assessing student conceptions with more conceptual complexity requires more rehearsals until a learning cycle is considered complete and before advancing to the next step. When learning to assess less complex student conceptions, one rehearsal appears to be sufficient. A noteworthy finding was that even after a second rehearsal phase, which in our study design consisted of student conceptions of a different content and different complexity, performance in assessing complex student conceptions was lower than the

performance of less complex student conceptions. Further research is needed to explore whether a second rehearsal assessing student conceptions with the same degree of complexity and the same content would further refine participants' assessment skills.

We point out that for the vignette test we used, no further test-theoretical quality criteria were available besides the reported reliability coefficients. Due to our specific subject-matter education context in which assessment skills were measured, we could not resort to fitting instruments with proven reliability and validity. In principle, however, the assessment of teachers' professional skills using video vignettes has proven successful many times (Hoppe, Renkl, Seidel, et al., 2020; Kersting, 2008; Schneider et al., 2016). Nevertheless, further studies should further investigate the validity of vignette-based assessment of preservice teachers' assessment skills in specific subject areas. Also, we recognize the significance of teachers' instructional decision-making subsequent to the assessment of student thinking. Blömeke et al. (2015) integrated teachers' decision-making into their competence model. We suggest that this model can guide further studies. Future research may investigate, for example, how preservice teachers can be supported in recomposing assessment of student thinking and instructional decision-making when enacting more complete practices of formative assessment (Grossman et al., 2009).

Conclusion

First, the study provides conceptual value because it outlines how on-the-fly assessment of student thinking about science can be promoted in preservice teachers through rehearsing category-based assessment of student thinking using video and written cases as representations of practice. Second, the study provides insight into how specifics of practice-based learning environments may be designed to effectively foster preservice teachers' on-the-fly assessment skills. Generally, teacher trainings using written and video cases as representations of practice, which provide participants with the opportunity to rehearse the assessment of student thinking, is effective. Written and video cases both serve as suitable representations of practice.

The conceptual complexity of the target content to be assessed, however, needs to be considered. Although a single rehearsal is effective for assessing less complex student conceptions, more complex content requires more practice. Given that the acquisition of assessment skills in student conceptions appears to be largely content-specific, participants always benefit from further rehearsals when a different content is presented. When sequences of rehearsals are being planned, different variations of media type (video or written cases) and content appear to have unique benefits in enhancing participants' assessment skills. Analyses of participants' immersion reveal a possible interaction between content complexity and media type on participants' engagement. We suggest considering a progression in the complexity of media type or content complexity when planning sequences of rehearsals.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the Ministry of Science, Research and Arts of Baden-Wuerttemberg (Research Training Group "Teachers' Diagnostic Competences") and the University of Education, Freiburg.

ORCID iD

Tobias Hoppe https://orcid.org/0000-0002-0149-8395

Note

1. In the literature on teacher training, the terms teacher *assessment* and teacher *diagnosing* are used with large conceptual overlap. This paper is relevant to both terms.

References

Blomberg, G., Stürmer, K., & Seidel, T. (2011). How pre-service teachers observe teaching on video: Effects of viewers' teaching subjects and the subject of the video. *Teaching and Teacher Education*, *27*(7), 1131–1140. https://doi.org/10.1016/j.tate.2011.04.008

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

Chi, M., Siler, S. A., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction*, *22*(3), 363–387. https://doi.org/10.1207/s1532690xci2203_4

Chi, M., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, *4*(1), 27–43. https://doi.org/10.1016/0959-4752(94)90017-5

Codreanu, E., Sommerhoff, D., Huber, S., Ufer, S., & Seidel, T. (2020). Between authenticity and cognitive demand: Finding a balance in designing a video-based simulation in the context of mathematics teacher education. *Teaching and Teacher Education*, *95*, Article 103146. https://doi.org/10.1016/j.tate.2020.103146

Danielson, K. A., Shaughnessy, M., & Jay, L. P. (2018). Use of representations in teacher education. In P. L. Grossman (Ed.), *Teaching core practices in teacher education* (pp. 15–34). Harvard Education Press.

Dannemann, S. (2020). Vorstellungen diagnostizieren [Diagnosing student conceptions]. In B. Reinisch, K. Helbig, & D. Krüger (Eds.), *Biologiedidaktische Vorstellungsforschung: Zukunftsweisende Praxis* (pp. 83–100). Springer.

Decristan, J., Hondrich, A. L., Büttner, G., Hertel, S., Klieme, E., Kunter, M., Lühken, A., Adl-Amini, K., Djakovic, S., S. K. Mannel, S., Naumann, A., & Hardy, I. (2015). Impact of additional guidance in science education on primary students' conceptual understanding.

The Journal of Educational Research, *108*(5), 358–370. https://doi.org/10.1080/00220671.2014.899957

Duit, R. (2009). *Bibliography—STCSE: Students' and Teachers' Conceptions and Science Education*. Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik. https://archiv.ipn.uni-kiel.de/stcse/

Duit, R., Treagust, D., & Widoro, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629–645). Routledge.

Farrell, M., Martin, M., Renkl, A., Rieß, W., Könings, K. D., van Merriënboer, J. J. G., & Seidel, T. (2022). An epistemic network approach to teacher students' professional vision in tutoring video analysis. *Frontiers in Education*, *7*, Article 805422. https://doi.org/10.3389/feduc.2022.805422

Fischer, J., Jansen, T., Möller, J., & Harms, U. (2021). Measuring biology trainee teachers' professional knowledge about evolution—Introducing the student inventory. *Evolution: Education and Outreach*, *14*(1), 1–16. https://doi.org/10.1186/s12052-021-00144-0

Friesen, M., & Kuntze, S. (2018). Competence assessment with representations of practice in text, comic and video format. In O. Buchbinder & S. Kuntze (Eds.), *Mathematics teachers engaging with representations of practice* (pp. 113–130). Springer. https://doi.org/10.1007/978-3-319-70594-1_7

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

Furtak, E. M., Thompson, J., & van Es, E. A. (2016, April 9). *Formative assessment and noticing: Toward a synthesized framework for attending and responding during instruction* [Paper presentation]. In Paper Presented at the Annual Meeting of the American Educational Research Association, Washington, DC, United States. https://doi.org/10.13140/RG.2.1.3471.1443

Gebhard, U., Höttecke, D., & Rehm, M. (2017). *Pädagogik der Naturwissenschaften* [Pedagogy of science]. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-19546-9

Gold, B., Hellermann, C., Burgula, K., & Holodynski, M. (2016). Fallbasierte Unterrichtsanalyse: Effekte von video- und textbasierter Fallanalyse auf kognitive Belastung, aufgabenspezifisches Interesse und die professionelle Unterrichtswahrnehmung von Grundschullehramtsstudierenden [Case-based analysis: Effects of video- and text-based case analysis on cognitive load, task-specific interest, and professional vision of preservice elementary school teachers]. *Unterrichtswissenschaft*, *44*(4), 322–338.

Gropengießer, H. (2006). *Wie man Vorstellungen der Lerner verstehen kann* [How to understand learners' ideas]. Didaktisches Zentrum Carl-von-Ossietzky-Univ.

Gropengießer, H., & Marohn, A. (2018). Schülervorstellungen und Conceptual Change [Student conceptions and conceptual change]. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (pp. 49–66). Springer.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*(9), 2055–2100.

Hammann, M., & Asshoff, R. (2015). *Schülervorstellungen im Biologieunterricht: Ursachen für Lernschwierigkeiten* [Student conceptions in biology classrooms: Causes for learning difficulties]. Klett Kallmeyer.

Hauser, M., & Kavanagh Schneider, S. (2019). Practice-based teacher education. In Noblit, G. W (Ed.), *Oxford research encyclopedia of education*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.419

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

Hoppe, T., Renkl, A., & Rieß, W. (2020). Förderung von unterrichtsbegleitendem Diagnostizieren von Schülervorstellungen durch Video- und Textvignetten [Fostering in-class diagnosis of student conceptions through video and text vignettes]. *Unterrichtswissenschaft*, *48*(4), 573–597. https://doi.org/10.1007/s42010-020-00075-7

Hoppe, T., Renkl, A., Seidel, T., Rettig, S., & Rieß, W. (2020). Exploring how teachers diagnose student conceptions about the cycle of matter. *Sustainability*, *12*(10), Article 4184. https://doi.org/10.3390/su12104184

Kavanagh, S. S., Metz, M., Hauser, M., Fogo, B., Taylor, M. W., & Carlson, J. (2020). Practicing responsiveness: Using approximations of teaching to develop teachers' responsiveness to students' ideas. *Journal of Teacher Education*, *71*(1), 94–107. https://doi.org/10.1177/0022487119841884

Kennedy, M. (2016). Parsing the practice of teaching. *Journal of Teacher Education*, *67*(1), 6–17. https://doi.org/10.1177/0022487115614617

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Education and Psychological Measurement*, *68*(5), 845–861.

Kloser, M. (2014). Identifying a core set of science teaching practices: A Delphi expert panel approach. *Journal of Research in Science Teaching*, *51*(9), 1185–1217. https://doi.org/10.1002/tea.21171

Kramer, C., König, J., Strauß, S., & Kaspar, K. (2020). Classroom videos or transcripts? A quasi-experimental study to assess the effects of media-based learning on pre-service teachers' situation-specific skills of classroom management. *International Journal of Educational Research*, 103, Article 101624. https://doi.org/10.1016/j.ijer.2020.101624

Loewenberg Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching. *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' diagnostic judgements by cognitive modeling (DiaCoM). *Teaching and Teacher Education*, 91, Article 103059.

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* [Qualitative content analysis: Basics and techniques]. Beltz.

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education. *Journal of Teacher Education*, *64*(5), 378–386. https://doi.org/10.1177/0022487113493807

Pine, K., Messer, D., & St John, K. (2001). Children's misconceptions in primary science: A survey of teachers' views. *Research*

in Science & Technological Education, *19*(1), 79–96. https://doi.org/10.1080/02635140120046240

Piwowar, V., Barth, V. L., Ophardt, D., & Thiel, F. (2017). Evidence-based scripted videos on handling student misbehavior: The development and evaluation of video cases for teacher education. *Professional Development in Education*, *10*(2), 1–16. https://doi.org/10.1080/19415257.2017.1316299

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, *37*(1), 15–24. https://doi.org/10.1016/j.stueduc.2011.04.003

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, *44*(1), 57–84. https://doi.org/10.1002/tea.20163

Scheiter, K. (2014). The learner control principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology: The Cambridge handbook of multimedia learning* (pp. 487–512). Cambridge University Press.

Schneider, J., Bohl, T., Kleinknecht, M., Rehm, M., Kuntze, S., & Syring, M. (2016). Unterricht analysieren und reflektieren mit unterschiedlichen Fallmedien: Ist Video wirklich besser als Text? [Analyzing and reflecting on teaching with different case media: Is video really better than text?]. *Unterrichtswissenschaft*, *44*(4), 474–490.

Schoenfeld, A. H. (2011). Noticing matters. A lot. Now what? In M. G. Sherin (Ed.), *Studies in mathematical thinking and learning. Mathematics teacher noticing: Seeing through teachers' eyes* (1st ed., pp. 223–238). Routledge.

Schrader, F. W., & Praetorius, A. K. (2018). Diagnostische Kompetenz von Eltern und Lehrern [Diagnostic competence of parents and teachers]. In D. H. Rost, J. R. Sparfeldt, & S. Buch (Eds.), *Handwörterbuch Pädagogische Psychologie* (5th ed., pp. 92–98). Beltz.

Seidel, T., Farrell, M., Martin, M., Rieß, W., & Renkl, A. (2022). Developing scripted video cases for teacher education: Creating evidence-based practice representations using mock ups. *Frontiers in Education*, 7, Article 965498. https://doi.org/10.3389/feduc.2022.965498

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education*, *27*(2), 259–267. https://doi.org/10.1016/j.tate.2010.08.009

Shavelson, R. J., Yue Yin Furtak, E. M., Ruiz-Primo, M. A., Ayala, C. C., Young, D. B., Tomita, M. K., Brandon, P. R., & Pottenger, F. M. III. (2008). On the role and impact of formative assessment on science inquiry teaching and learning. In J. E. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing science learning: Perspectives from research & practice* (pp. 21–36). National Science Teaching Association.

Shulman, L. S. (1984). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Stroupe, D., Hammerness, K., & McDonald, S. (2020). Practice-based science teacher education. In D. Stroupe, K. Hammerness, & S. McDonald (Eds.), *Preparing Science Teachers Through Practice-Based Teacher Education* (pp. 1–12). Harvard Education Press.

Syring, M., Bohl, T., Kleinknecht, M., Kuntze, S., Rehm, M., & Schneider, J. (2015). Videos oder Texte in der Lehrerbildung? [Videos or texts in teacher education?]. *Zeitschrift für Erziehungswissenschaft*, *18*(4), 667–685. https://doi.org/10.1007/s11618-015-0631-9

Taber, K. S. (2017). The nature of student conceptions. In K. S. Taber & B. Akpan (Eds.), *New directions in mathematics and science education: Volume 31. Science education: An international course companion* (pp. 119–131). Sense Publishers.

van Es, E. A. (2011). A framework for learning to notice student thinking. In M. G. Sherin (Ed.), *Studies in mathematical thinking and learning. Mathematics teacher noticing: Seeing through teachers' eyes* (1st ed., pp. 134–151). Routledge.

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. (2021). Expanding on prior conceptualizations of teacher noticing. *ZDM—Mathematics Education*, *53*(1), 17–27. https://doi.org/10.1007/s11858-020-01211-4

Windschitl, M., & Barton, A. (2016). Rigor and equity by design: Locating a set of core teaching practices for the science education community. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research* (pp. 1099–1158). American Educational Research Association. https://doi.org/10.3102/978-0-935302-48-6-18

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, *96*(5), 878–903. https://doi.org/10.1002/sce.21027

Author Biographies

Tobias Hoppe is a postdoctoral scholar at the University of Education Freiburg, Germany. His research is focused on explaining, assessing, and fostering teachers' formative assessment practices as well as the implementation of education for sustainable development in in-school and out-of-school learning environments. He previously taught high school in Southern Germany.

Tina Seidel is a professor of educational psychology at the TUM School of Social Sciences and Technology. Her research is focused on teacher professional vision, teacher-student interactions in classrooms, and evidence-based teacher education. She is particularly interested in developing and testing educational technology tools for higher education and, thereby, incorporating educational research on professional vision and classroom interactions in evidence-based learning environments.

Alexander Renkl is a professor of educational psychology and developmental psychology at the University of Freiburg. His main research areas are example-based learning, instructional explanations and self-explanations, learning from multiple representations, learning strategies, learning by journal writing, retrieval practice, and teachers' knowledge.

Werner Rieß is a professor of biology education, director of the Research Center for Climate Change Education and Education for Sustainable Development (ReCCE), and head of the Institute of Biology Education at the University of Education Freiburg. His research focuses on the impacts of education for sustainable development, and ways to measure and promote sustainability and climate literacy as well as systems thinking.