Chapter 4
Diagnosing Mathematical Argumentation Skills: A Video-Based Simulation for Pre-Service Teachers

Elias Codreanu, Sina Huber, Sarah Reinhold, Daniel Sommerhoff, Birgit J. Neuhaus, Ralf Schmidmaier, Stefan Ufer, and Tina Seidel

This chapter's simulation at a glance

Domain	Teacher education
Topic	Mathematical argumentation skills in the context of geometrical proofs
Learner's task	Taking on the role of pre-service interns and diagnosing the mathematical argumentation skills of four simulated seventh graders
Target group	Pre-service mathematic teachers
Diagnostic mode	Individual diagnosing
Sources of information	Interaction; observation of videos showing one-on-one student–teacher interactions
Special features	Simulated on-the-fly formative assessment situations

E. Codreanu (✉) · S. Huber
Department Educational Sciences, TUM School of Social Sciences and Technology, Technical University of Munich (TUM), Munich, Germany
e-mail: elias.codreanu@tum.de

S. Reinhold
Chair for Learning and Instruction, Center for International Student Assessment (ZIB), Technical University of Munich (TUM), Munich, Germany

D. Sommerhoff
Mathematics Education, IPN-Leibniz Institute for Science and Mathematics Education, Kiel, Germany

B. J. Neuhaus
Biology Education, Faculty for Biology, LMU Munich, Munich, Germany

R. Schmidmaier
Institute for Medical Education, University Hospital, LMU Munich, Munich, Germany

Department of Internal Medicine IV, University Hospital, LMU Munich, Munich, Germany

F. Fischer, A. Opitz (eds.), *Learning to Diagnose with Simulations*,
https://doi.org/10.1007/978-3-030-89147-3_4

4.1 Diagnosing Based on Student Observation

Every day, teachers face a variety of diagnostic situations in which they gather information about their students' learning prerequisites, processes, and outcomes (Herppich et al., 2018; Praetorius et al., 2013; Ruiz-Primo & Furtak, 2007; Thiede et al., 2015). This information serves as a basis for different pedagogical decisions like lesson planning, adaptive teaching, or grading students (Schrader, 2013; Dünnebier et al., 2009; Südkamp et al., 2012; Vogt & Rogalla, 2009). In particular, diagnostic decisions are indispensable for the continuous, on-the-fly adaptation of one's teaching to students' individual needs and ongoing learning processes. Across educational systems, such diagnostic situations arise within the everyday student–teacher interactions that dominate classrooms (Klug et al., 2013; Furtak et al., 2016; Kingston & Nash, 2011; Birenbaum et al., 2006). Teachers require professional vision to glean significant information from these classroom situations and reason about them (Seidel & Stürmer, 2014). During such high-density interactions, they engage in describing, evaluating, and explaining in order to make meaningful decisions about pedagogical actions.

For pre-service teachers, these high-density interactions are often experienced as overwhelming, since they require the deliberate employment of diagnostic decision-making (Levin et al., 2009). Therefore, many pre-service teachers struggle to find their way around into the profession (Stokking et al., 2003). Nevertheless, diagnostic skills for diagnostic situations in the classroom are rarely taught in teacher education. Initially, university teacher education focuses on conveying basic principles and conceptual knowledge, often separated into different fields related to content knowledge, pedagogical content knowledge, and educational psychology. Given these structures, it is often unclear how these aspects of professional knowledge are related to specific diagnostic situations in classrooms (Alles et al., 2019). Therefore, new ways of supporting the acquisition of crucial skills like diagnostic skills are needed to prepare pre-service teachers to make reasonable diagnostic decisions before they enter their first classroom. Additionally, little is known about the processes involved in diagnostic decision-making and differences in these processes along the learning trajectory (Herppich et al., 2018). Insights into these processes may be promising to identify characteristics for targeted interventions along this learning trajectory.

S. Ufer
Chair of Mathematics Education, LMU Munich, Munich, Germany

T. Seidel
Friedl Schöller Endowed Chair for Educational Psychology, School of Education, Technical University of Munich (TUM), Munich, Germany

4.2 Simulation as a Model of Reality

An environment to investigate and promote pre-service teachers' diagnostic skills should encompass two aspects: First, following a practice-oriented approach, it should represent practice in an authentic way in order to motivate pre-service teachers to get involved in the actual task (Schubert et al., 2001). This, in turn, allows pre-service teachers to transfer their behavior from the simulated environment to real-world teaching situations. Second, reality must be decomposed and simplified in a way that enables pre-service teachers to focus on particular aspects of classroom situations (Grossman et al., 2009). Such decompositions of practice contain key features that make diagnostic decision-making more accessible to pre-service teachers than in real-world classroom situations.

Due to its strengths in both respects, video is becoming a frequently used medium in professional teacher education (Kang & van Es, 2018; Gaudin & Chaliès, 2015). Although videos can capture only one perspective on a classroom situation, and thus have a limited ability to convey the contextual background of the situation, videos can give authentic insights into different teaching and learning situations (Blomberg et al., 2013). Moreover, by taking a certain perspective, videos can direct observers' attention to significant features of the situation using so-called cues. Applying the idea of decomposing practice, in the specific sense of diagnosing students' skills based on observing them in the classroom, videos should capture everyday student-teacher interactions, including the most relevant cues for diagnosis but only a few less relevant cues that can distract teachers' attention in real-world classrooms. If the goal is to diagnose mathematical argumentation skills from a mathematics educational perspective, the most relevant cues include students' statements regarding their understanding of correct mathematical proof procedures, for example. General aspects like students' situational motivation can be considered less relevant for making such diagnoses. Reducing the number of less relevant cues increases teachers' capacity for deliberate action. The scripted video format also uniquely allows for further targeted manipulation of these segments (Piwowar et al., 2017).

Not just the makeup of scripted videos but also their embedding in a simulated environment influences learning grounded in practice. Decomposing the situation by dividing a scripted video into a number of scenes provides an opportunity to slow down the actual situation and thereby reduce the density of interactions. By decomposing situations, simulations provide researchers with insights into processes and allow for gathering data for further analyses of diagnostic skills. The results of such analyses may then help to develop future evidence-based interventions.

4.3 Mathematical Argumentation Skills

Mathematics is a relevant subject for studying diagnostic situations involving student–teacher interactions because the traditional initiation-response-feedback teaching discourse is the prevalent form of teacher–student dialogue (Lipowsky et al., 2009). In mathematics, as a proof-based science, working with mathematical argumentation as well as with proofs, as a special form of this argumentation fulfilling strict standards (Stylianides, 2007), is a crucial learning activity. Mastery of these skills is a central learning goal in many secondary school systems (Kultusministerkonferenz., 2012). However, empirical studies have repeatedly shown that students have substantial problems when attempting to construct a mathematical proof (Healy & Hoyles, 2000; Harel & Sowder, 1998). In particular, being able to successfully construct mathematical proofs depends on different individual prerequisites (Sommerhoff et al., 2015; Schoenfeld, 1992). These factors can be used in the diagnostic situation as indicators for diagnosing students' skills in working with argumentations and proofs. Ufer et al. (2008) and Sommerhoff et al. (2015) emphasize students' *mathematical content knowledge*, *methodological knowledge*, and *problem-solving strategies* as three important prerequisites. However, these three prerequisites can be divided into more specific components for use in the diagnostic process, as described in the following paragraph.

Mathematical content knowledge comprises three different sub-concepts (Weigand et al., 2014). First, knowledge of *concept properties* encompasses knowledge of features and terms, like features of the diagonals of parallelograms. The second sub-concept, known as *concept scope*, concerns knowledge of the entirety of representatives of a mathematical term. For example, this includes the knowledge that a square is a representative of the term parallelogram. Third, the *concept network* refers to knowledge about the relationship between a concept and other concepts. Likewise, methodological knowledge—that is, knowledge about the nature of proofs, their use within mathematics, and socio-mathematical norms regarding proofs—can be divided into at least three components (Heinze & Reiss, 2003): Knowledge of *proof scheme* encompasses knowledge about acceptable types of inferences in a proof. *Proof structure*, in contrast, refers to the overall logical structure of a proof, such as starting with assumptions and ending with an assertation. Finally, *chain of conclusion* refers to the logical arrangement of individual arguments within the proof. With respect to problem-solving strategies, this research project focuses on two different aspects. First, *heuristic strategies* help to solve a given problem task by reorganizing the task and changing how one looks at it. Second, *metacognitive strategies* allow an individual to control the problem-solving process through strategies such as monitoring and assessing their progress within the problem-solving process and drawing conclusions for action.

Prior research indicates that students typically differ widely regarding each of these eight aspects, resulting in a range of difficulties when attempting mathematical proofs (Reiss & Ufer, 2009). It is a difficult task for teachers to diagnose the reasons for students' difficulties and thus also what form of teacher support will help each individual student based solely on brief student–teacher interactions and possibly a brief look at students' notes.

4.4 Guiding Questions in Designing the Simulation

Both measuring and supporting teachers' diagnostic skills via simulations require high standards in terms of the simulations' authenticity and the content of the embedded videos. The development of the video-based simulation presented in this chapter was thus guided by the following questions (Codreanu et al., 2020):

1. To what extent can we authentically represent a diagnostic situation within student–teacher interaction around mathematical argumentation in a scripted video-based simulation?
2. To what extent can the decomposition of the diagnostic situation in the video-based simulation provide insights into the participants' diagnostic processes?

4.5 Conceptualization of the Scripted Videos

To create a simulated setting for diagnosing individual students' mathematical argumentation skills in a simulated classroom situation, we developed scripted videos with small-group student–teacher interactions, following Dieker et al. (2009)'s recommendations. First, we identified the essential features of the relevant situation (*selection of practice*). Second, we developed a contextual frame for all of the recordings as well as detailed scripts for each scene (*vignette script development*). Third, we created the video footage and edited it to create a representation of teaching practice (*video production*).

Selection of Practice We decided to focus on three individual student prerequisites that are important predictors of their performance when working with proofs (Ufer et al., 2008): (a) *mathematical content knowledge*, (b) *methodological knowledge*, and (c) *problem-solving strategies*. All three prerequisites have been shown to affect students' skills in working with geometrical proofs and can be portrayed in brief video clips. We considered the three sub-concepts of mathematical content knowledge, the three sub-concepts of methodological knowledge, and the two aspects of problem-solving strategies as a theoretical fundament when designing the student profiles.

Afterwards, we outlined four student profiles varying in their levels of the aforementioned prerequisites of students' skills in working with argumentations and proofs (eight aspects in total). Van Hiele's model of children's development of geometric thinking provided valuable additional guidance in this context (Usiskin, 1982). According to this model, a student on the first level can recognize and judge

	A	B	C	D
Mathematical Content Knowledge				
Concept Properties	★	★ ★	★ ★ ★ ★	★ ★ ★ ★
Concept Scope	★	★ ★	★ ★ ★ ★	★ ★ ★ ★
Concept Network	★	★ ★	⃠	★ ★ ★ ★
Methodological Knowledge				
Proof Scheme	★	★	★ ★ ★	★ ★ ★ ★
Proof Structure	★	★ ★	★ ★ ★	★ ★ ★
Chain of Conclusion	★	★	★ ★	★ ★ ★
Problem-Solving Strategies				
Heuristic Strategies	★ ★ ★	★ ★ ★ ★	★ ★ ★ ★	★ ★ ★ ★
Metacognitive Strategies	★	★ ★	★ ★ ★	★ ★ ★ ★

⃠	not observable	★	mostly not available	★★	insufficiently available	★★★	mostly available	★★★★	available

Fig. 4.1 Four student profiles and their specific predictive prerequisites for performance in working with proofs

figures by their appearance. A student on the second level can identify properties of figures, while a student on the third level can already follow simple deductions. Only at level four has a student acquired sufficient understanding to meaningfully construct proofs. We concentrated on these four levels and specified the student profiles based on their knowledge and abilities with respect to the eight predictive aspects. For example, Profiles A and B know little about what inferences are acceptable (*proof scheme*), whereas Profile C by and large and Profile D fully understand this point (see Fig. 4.1).

To ensure that the simulated students remained comparable, all simulated students worked on the same geometry proof task in the video clips: They had to prove that opposite sides of a parallelogram are of equal length, based on the information that pairs of sides of a parallelogram are parallel. Students who are just beginning to learn how to work with proofs do not pay a lot of attention to norms and standards of proofs on an abstract level. Thus, we did not expect all aspects, especially those for methodological knowledge, to become important in the proof construction process for these students. Likewise, it is possible that not all four simulated students need to use the conceptual network during the proof construction process. This is why we took care to select a task that can be completed in different ways to serve as a basis for the simulation.

Vignette Script Development for Staged Videos The time the simulated students spent working on the geometrical proof task was split into eight smaller video scenes, each with a length of approximately 1 min. Thus, all simulated students were depicted in the same number of scenes, which was sufficient to provide participants with the opportunity to observe each simulated student multiple times. The scripts for these scenes contained detailed dialogues between the teacher and simulated student, as well as copies of the simulated students' sketches and other notes in their exercise books. The teacher's input in the scenes was reduced to a minimum, focusing solely on eliciting the simulated students to talk about their thoughts. Thus, typical questions and requests by the teacher were "What do you mean by that?" or "Can you explain what you have done here?"

The answers and statements given by the simulated students were generated according to their profiles and under consideration of the eight identified aspects of predictive prerequisites. Cues could be found not only in the verbal teacher-student interaction but also in the students' sketches and notes. When creating the scenes, the cue attributions were continuously reviewed in an internal review process to ensure that the video scenes provided salient cues for the prerequisites. These cues were distributed as evenly as possible over the eight scenes in order to portray an authentic conversation. This resulted in a distribution in which at least one (and often more than one) salient cue for each aspect occurred no later than the fourth video scene.

Production of Staged Videos The video-recording was completed with one trained teacher and four eighth-grade student volunteers. The teacher and students were provided with the scripts prior to filming and were given time and guidance to familiarize themselves with their role, the script, and each other. While the scene between the teacher and one student was being filmed, the other students practiced their next scene with a member of the video production team. During shooting, the actors followed the scripts with as much fidelity as natural behavior allowed in that moment. The research team ensured that the main cues within the scripts were successfully captured on video. To capture both the verbal student–teacher interaction and the students' written notes, two different camera perspectives were used at the same time: One from the front showing the conversation, and one from above showing the student's exercise book. In the editing process, the scenes were cut to show the appropriate camera angle in each moment. After production was complete, the final video scenes were reviewed by two independent researchers with respect to the perceptibility of the cues contained in the initial scripts. In a subsequent consensus process based on the final video scenes, the four student profiles were classified into four ordinal categories with respect to each predictive prerequisite (see Fig. 4.1).

4.6 Design of the Simulation

The presented video-based simulation has an underlying structure consisting of four main parts (see Fig. 4.2). It starts by familiarizing participants with the situation depicted in the simulation, a pre-service intern observing student-teacher interactions during a student exercise concerning a geometric proof (*diagnostic situation*). Before participants start working with the tool, they are introduced to the task to be accomplished in this diagnostic situation: the simulated teacher asks them to assess the students' mathematical argumentation skills so that he can choose tasks for individual learning support in a subsequent class based on the participants' observations (*diagnostic task*). After that, participants can work independently in the simulated classroom situation to gather information about the students by watching the video scenes and taking notes (*diagnostic process*). These notes form the basis for the final diagnosis of each simulated student, which participants formulate in the last section of the tool to provide the simulated teacher a basis for his further lesson planning (*diagnostic outcome*).

Diagnostic Situation The situation chosen for this simulation is an everyday classroom situation in mathematics lessons (Lipowsky et al., 2009). Students are working independently on a task, in this case, a geometry proof, while the teacher walks from student to student to monitor and support their progress in short student-teacher interactions. At the beginning of the simulation, participants are familiarized with their role in this simulation: they are observing the teacher and students' interactions in their role as pre-service intern. In addition, they receive information about the overall topic, prior lessons, and learning context in order to acquaint them with the classroom situation as well as with the content discussed in the lesson. Taking on the role of an intern is familiar to participants (pre-service teachers), so they should be able to put themselves in this role without all too much effort. Thus, the scenario is likely to support immersion into the simulation (Slater & Wilbur, 1997). Furthermore, interns in real-life classrooms face similar challenges and opportunities to the ones contained in the diagnostic process later in the simulated situation. This parallelism between an intern's role in real-world situations and in the simulated diagnostic situation is expected to lead to higher authenticity of the learning environment (Schubert et al., 2001).

After the introduction to their role, participants receive information about the different steps a teacher considers when preparing a lesson. Information about the prior knowledge of the whole class and the topics covered in the class's previous lessons is provided. In addition, participants have an opportunity to familiarize themselves with the proof task for the upcoming lesson.

Diagnostic Task After familiarizing themselves with the diagnostic situation, the simulated teacher presents the diagnostic task to the participants. They are asked to diagnose four specific simulated students' level of understanding of working with geometric proofs in order to give the simulated teacher ideas for individual student support in a subsequent remedial lesson.

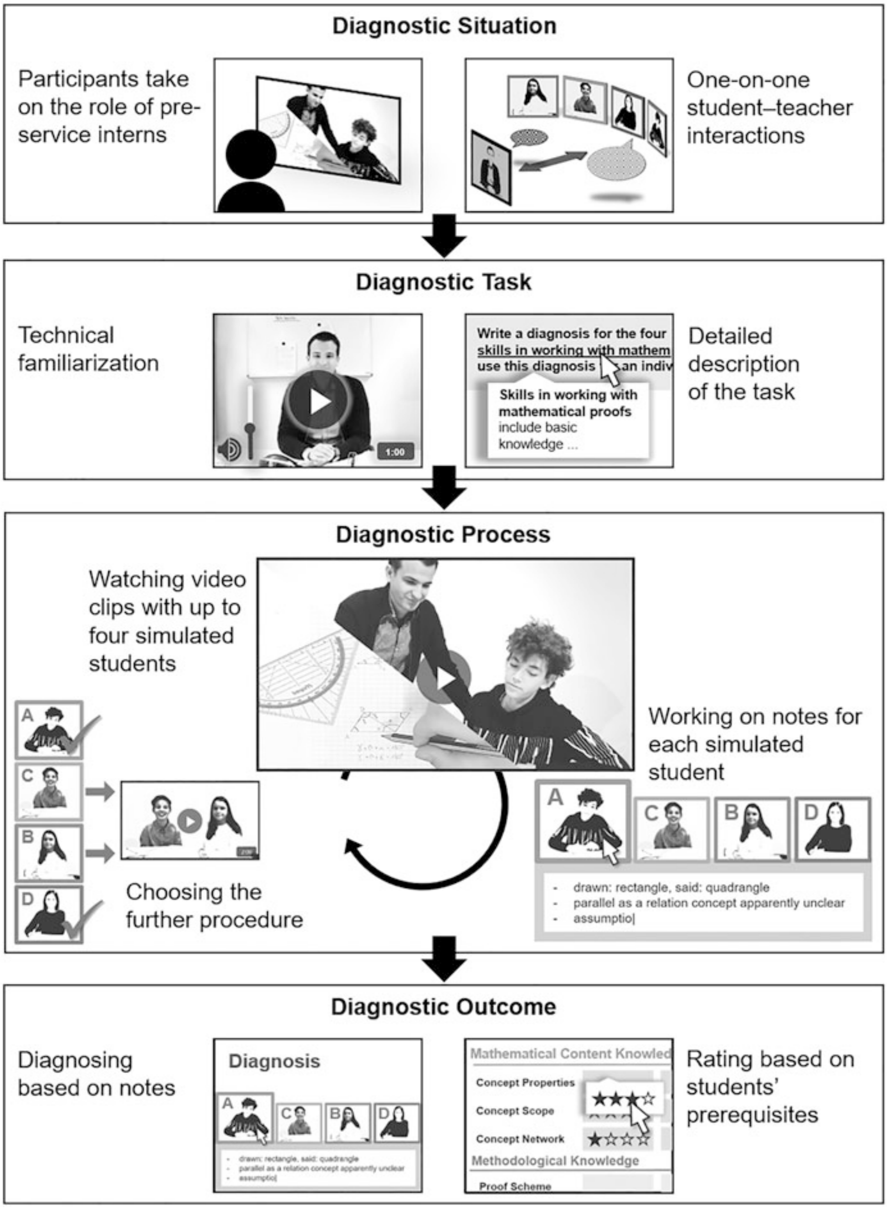

Fig. 4.2 Design of the video-based simulation. Note: Adapted from "Between authenticity and cognitive demand: Finding a balance in designing a video-based simulation in the context of mathematics teacher education" by Codreanu et al., 2020, *Teaching and Teacher Education*, 95, 103,146

We strive for two goals during the presentation of the diagnostic task: participants should come to understand both the specific task in detail and its embedding in the simulation. On the one hand, presenting the task during a short video clip familiarizes participants with the technical aspects of the simulation. For example, participants have the possibility to play and pause but not rewind the video in order to more closely simulate reality. This technical familiarization aims to minimize technical complications later in the diagnostic process. Additionally, participants get to know the teacher they will accompanying in the subsequent simulation. As a second major aim, the diagnostic task is described in detail, focusing on the following two aspects: (a) *who* is the diagnosis for and thus *how* should it look, and (b) what is the diagnosis' *purpose* and which *components* should it therefore entail? Considering that the participants most likely have little experience with diagnosing students' skills and abilities and the terminology used in this field, we provide a detailed description of the task to be completed in the subsequent diagnostic process. Regarding aspect (a), it is pointed out that a diagnosis should include descriptions, explanations, and decisions (Blömeke et al., 2015; Seidel & Stürmer, 2014). Regarding aspect (b), a description of the ability to work with geometrical proofs is provided, addressing the predictive prerequisites implemented in the video clips (see Fig. 4.1).

Diagnostic Process During the diagnostic process, the participants observe four simulated students, which simulates a reduced classroom setting. The process is divided into several cycles. Each cycle starts with watching one video clip containing student–teacher interaction scenes between one of the simulated students and the teacher. In the first cycle, participants observe all four simulated students in a row. Participants can take notes while observing the simulated students. Participants can enter their notes for each simulated student in the respective text box by clicking on the picture of each simulated student. This makes it possible to take individual notes for each simulated student. At the end of a cycle, participants must choose whether or not to continue the procedure. They can decide whether they want to observe more interactions with each student and thus run through another cycle of the diagnostic process for them, or conclude the diagnostic process for this particular simulated student. Thus, if a participant decides to continue observing two of the four simulated students, for example, the next cycle shows only these two students' further work on the proof task. Only the text boxes for the two remaining simulated students can be opened. After this second cycle has been completed, participants again decide whether to continue to observe each of the remaining simulated students in a third cycle. This continues until participants choose to conclude the observation process for all four simulated students or after a maximum of eight cycles.

In the first cycle, participants start with an empty text box for their note-taking. In subsequent cycles, notes from the previous cycles are already displayed in the text box, so that participants can further add to their previous notes. These notes serve as individual support to participants throughout the entire diagnostic process. However, the maximum number of scenes participants can watch is limited to 20. Thus, they must allocate the number of scenes they watch depicting each of the four simulated students. This also makes it possible to measure the efficiency of the participants' diagnostic process.

Diagnostic Outcome Finally, after participants complete the diagnostic process for all simulated students, they have to submit their diagnoses in two different ways. First, they are asked to formulate a diagnosis for each simulated student in an open-response text box. Their notes from the diagnostic process are shown above the text box. The participants can copy parts of the notes, summarize their points, or use the notes as an aid to remember the situations in the video clips. Like in the notes page, they can work on the four diagnoses in any order. Second, participants are asked to assess the simulated students' mathematical content knowledge, methodological knowledge, and problem-solving strategies. Participants have to rate the students' possession of each of the eight predictive prerequisites on a four-point Likert scale. These two tasks allow for participants' diagnostic outcomes to be assessed in two different ways, enabling a more differentiated consideration (see Fig. 4.1). Additionally, participants are asked to rank the student profiles according to their level of mathematical argumentation skills from weakest to strongest.

4.7 Discussion and Outlook

The video-based simulation developed in this project provides an innovative way to investigate and promote pre-service teachers' diagnostic skills regarding students' mathematical argumentation skills. The described development process is likely crucial for the effectiveness of video-based simulations targeting diagnostic skills in teacher education (see overarching research question 1 in Fischer et al., 2022). The purposeful conceptualization of the scripted videos and the simulation design suggest that the environment represents practice authentically and allows participants to immerse themselves in the situation. This supports the transfer of behavior to real-world situations. The specific facet of practice chosen for the scripted videos, namely the geometry task and student-teacher interactions surrounding it, resemble situations found in real-world mathematics classrooms. Moreover, the four meticulously designed student profiles capture important student prerequisites in terms of mathematical content knowledge, methodological knowledge, and problem-solving strategies (Ufer et al., 2008). Finally, the video clips were filmed with student volunteers, who enriched the script with their natural behavior. In the simulation itself, we separated the content-related and technical familiarization with the task from the part where the participants actually work on the simulation task. Hence, all information required to work undisturbed on the task and all additional instructions on the simulated situation are provided before the actual diagnostic process starts. This makes it possible to immerse oneself more deeply into the situation. In empirical analyses, expert teachers' and pre-service teachers' ratings of the authenticity and immersion of the scripted videos and the simulation as a whole are used to evaluate whether participants experience the simulated learning environment as a convincing representation of real-world classrooms (e.g. Codreanu et al., 2020). We involve expert teachers due to their wealth of experience in classroom situations, and novice teachers because they represent the target group for whom the simulation was developed. These and other variables are likely moderating and mediating variables for the successful completion of the simulation as well as embedding additional instructional support in the simulation (see overarching research question 3 in Fischer et al., 2022).

The specific conceptualization and design of the scripted videos and their embedding in the simulation both contribute to decomposing practice in a way that allows for the extraction of features regarding the participants' diagnostic process (see overarching research question 4 in Fischer et al., 2022). The scripted videos depict only four simulated students whose profiles differ only with regard to important prerequisites for successfully completing geometrical proofs. This makes it easier for participants to focus on and distinguish between students than in a classroom with twenty-plus students with more diverse compositions of those prerequisites. The deliberate absence of classroom management issues such as handling disturbances gives participants the opportunity to concentrate on more relevant rather than less relevant cues in their diagnostic processes. Adding time to the participants' observations by having them take notes slows down the ongoing classroom actions. While a real-world classroom does not include specific times to take notes on what teachers notice and interpret, the simulation does include these processes. Furthermore, the instructions to both describe and interpret one's observations in the notes helps teacher process in detail what they have observed. This reduces the complexity of the situation and allows participants to record important mental steps. Additionally, the notes give insight into participants' reasoning use and performance (Seidel & Stürmer, 2014). These data can help identify key features in the diagnostic process in order to develop targeted support within the simulation. Thus, analyzing pre-service teachers' diagnostic processes should reveal differences in where instructional support like scaffolding and prompts can be set (see overarching research question 2 in Fischer et al., 2022).

We expect to obtain further findings on the processes and variables that influence simulation performance by investigating participants' individual prerequisites, like their knowledge base or interest and self-concept. Based on these findings, the simulation will be expanded from a tool to assess diagnostic skills into a tool that is also able to foster those skills.

Acknowledgments The research presented in this chapter was funded by a grant from the Deutsche Forschungsgemeinschaft (DFG-FOR 2385) to Tina Seidel, Stefan Ufer, Birgit Neuhaus, and Ralf Schmidmaier (SE 1397/11-1).

References

Alles, M., Apel, J., Seidel, T., & Stürmer, K. (2019). How candidate teachers experience coherence in university education and teacher induction: The influence of perceived professional preparation at university and support during teacher induction. *Vocations and Learning, 12*, 87–112. https://doi.org/10.1007/s12186-018-9211-5

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., et al. (2006). A learning integrated assessment system. *Educational Research Review, 1*, 61–67. https://doi.org/10.1016/j.edurev.2006.01.001

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

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

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*, 103146. https://doi.org/10.1016/j.tate.2020.103146

Dieker, L. A., Lane, H. B., Allsopp, D. H., O'Brien, C., Butler, T. W., Kyger, M., et al. (2009). Evaluating video models of evidence-based instructional practices to enhance teacher learning. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 32*, 180–196. https://doi.org/10.1177/0888406409334202

Dünnebier, K., Gräsel, C., & Krolak-Schwerdt, S. (2009). Urteilsverzerrungen in der schulischen Leistungsbeurteilung. *Zeitschrift für Pädagogische Psychologie, 23*, 187–195. https://doi.org/10.1024/1010-0652.23.34.187

Fischer, F., Chernikova, O., & Opitz, A. (2022). Learning to diagnose with simulations: Introduction. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations – examples from teacher education and medical education*. Springer.

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

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

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

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *CBMS Issues in Mathematics Education, 7*, 234–283.

Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*(4), 396–428.

Heinze, A., & Reiss, K. (2003). Reasoning and proof: methodological knowledge as a component of proof competence. www.lettredelapreuve.itlCERME3PapersiHeinze-paperl.pdf.

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., et al. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education, 76*, 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Kang, H., & van Es, E. A. (2018). Articulating design principles for productive use of video in preservice education. *Journal of Teacher Education, 5*, 1–14. https://doi.org/10.1177/0022487118778549

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*, 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

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

Kultusministerkonferenz. (2012). *Bildungsstandards im Fach Mathematik für die allgemeine Hochschulreife*. KMK.

Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. *Journal of Teacher Education, 60*, 142–154. https://doi.org/10.1177/0022487108330245

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

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, 44*, 369–384. https://doi.org/10.1080/19415257.2017.1316299

Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *The Journal of Educational Research, 106*, 64–76. https://doi.org/10.1080/00220671.2012.667010

Reiss, K., & Ufer, S. (2009). Was macht mathematisches Arbeiten aus? Empirische Ergebnisse zum Argumentieren, Begründen und Beweisen. *Jahresbericht der Deutschen Mathematiker-Vereinigung, 111*(4), 1–23.

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*, 57–84. https://doi.org/10.1002/tea.20163

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacogntion, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). Simon & Schuster.

Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerbildung, 31*(2), 154–165.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence, 10*(3), 266–281.

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

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 6*(6), 603–616.

Sommerhoff, D., Ufer, S., & Kollar, I. (2015). Research on mathematical argumentation: A descriptive review of PME proceedings. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of 39th Psychology of Mathematics Education conference* (pp. 193–200). PME.

Stokking, K., Leenders, F., de Jong, J., & van Tartwijk, J. (2003). From student to teacher: Reducing practice shock and early dropout in the teaching profession. *European Journal of Teacher Education, 26*, 329–350. https://doi.org/10.1080/0261976032000128175

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289–321.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*, 743–762. https://doi.org/10.1037/a0027627

Südkamp, A., & Praetorius, A.-K. (2017). *Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen (Pädagogische Psychologie und Entwicklungspsychologie)* (Vol. 94). Waxmann.

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., et al. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education, 49*, 36–44. https://doi.org/10.1016/j.tate.2015.01.012

Ufer, S., Heinze, A., & Reiss, K. (2008). Individual predictors of geometrical proof competence. *PME 32 and PME-NA XXX, 1*(4), 361–368.

Usiskin, Z. (1982). Van Hiele levels and achievement in secondary school geometry: CDASSG Project.

Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education, 25*, 1051–1060. https://doi.org/10.1016/j.tate.2009.04.002

Weigand, H.-G., Filler, A., Hölzl, R., Kuntze, S., Ludwig, M., Roth, J., et al. (2014). Didaktik der Geometrie für die Sekundarstufe I. In *Mathematik Primarstufe und Sekundarstufe I + II* (2nd ed.). Springer Spektrum.