Technische Universität München
TUM School of Social Sciences and Technology

Developing Effective Scaffolding for Future Mathematics Teachers' Assessment Skills in Simulations: Towards Adaptivity

Michael Ludwig Nickl

Vollständiger Abdruck der von der TUM School of Social Sciences and Technology der Technischen Universität München zur Erlangung eines

Doktors der Philosophie (Dr. phil.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Tilman Michaeli

Prüfer*innen der Dissertation:

1. Prof. Dr. Christina Seidel
2. Prof. Dr. Andreas Obersteiner

Die Dissertation wurde am 05.04.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Social Sciences and Technology am 10.06.2023 angenommen.

Acknowledgements

This dissertation would not have been possible without the support of many people.

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Dr. Tina Seidel, for giving me the wings to explore the fascinating world of research. She provided support whenever necessary, and she let me explore the world by myself whenever possible. With her kind and well-informed support, she helped me to develop and improve my research skills, and with her enthusiasm and research experience, she inspired me to pursue my own research endeavor. I also want to thank my mentor, Prof. Dr. Daniel Sommerhoff, who was available all day and night to support me in the daily business of research and for keeping me on track. His continuous support motivated me, and the many talks we had sharpened my research goals. Moreover, I want to thank Prof. Dr. Andreas Obersteiner for joining the examination committee, and Prof. Dr. Tilman Michaeli for chairing it. I also want to thank Prof. Dr. Jan Plass for giving me the opportunity to do a research stay at New York University. It was a pleasure to benefit from his great experience and to widen my perspective on research through all the amazing experiences I had at NYU. Moreover, I want to thank the Visit-Math project team, including the student assistants, for all the support and valuable feedback I received. The joint collaboration for the sake of designing an innovative learning environment and conducting research about it was more than intriguing. This project work was complemented by the many colleagues at the Friedl Schöller Endowed Chair for Educational Psychology at Technische Universität München (TUM) that I may call friends. It is a great pleasure working with you, and I can remember lots of fun times with you – on-task and off-task. Many thanks also go to the COSIMA research group. Seeing people from different disciplines with different perspectives working together to obtain a deeper understanding of a research topic was intriguing, and I am glad to be a part of it.

I also want to thank Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding this research (Grant number 1397/11-2, FOR 2385).

Finally, I want to express my sincere gratitude to all of my friends and to my family for accompanying me on my journey of this dissertation, for encouraging me to continue on my path, and for providing diversion whenever I needed it, with all their empathy, kindness, and love.

Abstract

As an important aspect of their professional practice, teachers have to accurately assess learning-relevant student characteristics, such as prior knowledge, for which they need assessment skills. However, as prior research suggests that teachers' assessment skills need improvement, facilitating their assessment skills in university teacher education seems promising, for example, by using video-based simulations. To further increase learning gains in such learning environments, the present dissertation developed, evaluated, and compared possible effective scaffolding strategies within a validated video-based simulation. To this end, two empirical studies with $N = 150$ and $N = 108$ future teachers were conducted. In the first study, a latent profile analysis was used to identify three typical patterns of future teachers' learner characteristics: future teachers were either particularly *knowledgeable*, particularly *motivated,* or had below-average knowledge and motivational characteristics. Future teachers of different learner characteristic patterns differed in their perceptions of situative learning experiences and in their assessment skills, thus revealing the potential for differentiated support. Based on these results, conceptual prompts as cognitive support, a utility value intervention as motivational support, and their combination were derived as promising scaffolding strategies and were empirically evaluated in the second study. The results showed that conceptual prompts increased judgment accuracy, particularly for future teachers with comparably low success expectancy. Future teachers with comparably high success expectancy benefitted the most from a utility value intervention. Surprisingly, the combination of both scaffolds resulted in the least learning gains compared to both scaffolds individually and the control group. These studies contribute to a deeper understanding of teachers' assessment skills by highlighting the important role of knowledge and, tentatively, the role of motivational characteristics in judgment accuracy. They also contribute to a deeper understanding of the use of scaffolding in simulations by substantiating the effectiveness of conceptual prompts and by pointing toward a varying effectiveness of conceptual prompts and the utility value intervention for learners with a comparably low/high success expectancy. Taken together, both studies provide an approach for developing effective scaffolding, which also paves the way for further adaptive support within simulations.

Zusammenfassung

Ein wesentlicher Aspekt der beruflichen Praxis von Lehrkräften ist die akkurate Einschätzung lernrelevanter Merkmale von SchülerInnen wie Vorwissen. Dafür benötigen Lehrkräfte Diagnosekompetenzen, die bisheriger Forschung zufolge verbesserungsbedürftig sind. Eine frühzeitige Förderung in der universitären Ausbildung von LehramtskandidatInnen durch videobasierte Simulationen scheint vielversprechend. Um Lernzuwächse in solchen Lernumgebungen weiter zu steigern, wurden in der vorliegenden Dissertation effektive Scaffolding-Strategien in einer bereits validierten videobasierten Simulation entwickelt, evaluiert und verglichen. Hierzu wurden zwei empirische Studien mit $N = 150$ und $N = 108$ Lehramtsstudierenden durchgeführt. In der ersten Studie wurden mittels latenter Profilanalyse drei Lernvoraussetzungsmuster von Lehramtsstudierenden identifiziert: Lehramtsstudierende verfügten entweder über überdurchschnittliches Wissen oder über überdurchschnittliche motivationale Charakteristika oder waren entlang beider Dimensionen unterdurchschnittlich. Lehramtsstudierende mit unterschiedlichen Mustern unterschieden sich in situationaler Wahrnehmung der Simulation und in Diagnosekompetenzen. Darauf aufbauend wurden konzeptuelle Prompts als kognitive Unterstützung, eine Utility Value Intervention als motivationale Unterstützung sowie deren Kombination als potentiell erfolgsversprechende Scaffolding-Strategien abgeleitet, die dann in der zweiten Studie empirisch evaluiert wurden. Dabei zeigte sich, dass die verwendeten konzeptuellen Prompts der diagnostischen Performanz zuträglich waren, insbesondere für Lehramtsstudierende mit relativ geringer Erfolgszuversicht. Lehramtsstudierende mit relativ hoher Erfolgszuversicht profitierten hingegen am ehesten von der Utility Value Intervention. Die Kombination beider Interventionen hatte überraschend die geringsten Lernzuwächse im Vergleich zu den einzelnen Interventionen und einer Kontrollgruppe. Für sich genommen, tragen die beiden Studien zu einem besseren Verständnis von Diagnosekompetenzen von Lehrkräften hinsichtlich der zentralen Rolle professionellen Wissens und der sich andeutenden Rolle motivationaler Charakteristika für die Diagnoseakkuratheit bei. Weiter tragen die Studien zum Verständnis von Scaffolding in Simulationen bei, insofern als die Rolle konzeptueller Prompts hervorgehoben und die differentielle Wirksamkeit von konzeptuellen Prompts und der Utility Value Intervention für Lehramtsstudierende mit relativ niedriger/hoher Erfolgserwartung angedeutet wurde. Zusammengenommen, zeigen die Studien einen Ansatz zur Entwicklung effektiver Scaffolding Strategien auf, der auch den Weg für adaptive Unterstützung in Simulationen ebnet.

Included Publications

The present dissertation is publication-based and consists of two articles published/accepted for publication in international peer-reviewed journals. The author of this dissertation is the first author of both publications and had the main responsibility for conceptualization, data analyses, writing the original and revised drafts, and the publication process at the respective journals.

Article A

Nickl, M., Huber, S. A., Sommerhoff, D., Codreanu, E., Ufer, S., & Seidel, T. (2022). Video-based simulations in teacher education: The role of learner characteristics as capacities for positive learning experiences and high performance. *International Journal of Educational Technology in Higher Education*, *19*, Article 45. https://doi.org/10.1186/s41239-022-00351-9

Article B

Nickl, M., Sommerhoff, D., Böheim, R., Ufer, S., & Seidel, T. (2023). Fostering pre-service teachers' assessment skills in a video simulation: Differential effects of a utility value intervention and conceptual knowledge prompts. Zeitschrift für Pädagogische Psychologie. https://doi.org/10.1024/1010-0652/a000362

Contents

1 Introduction

In their professional practice, teachers are confronted with multiple situations that require the assessment of students. For example, teachers need to assess students' relevant prior knowledge when preparing a learning sequence, they need to assess students' learning activities during lessons, and they may assess learning gains at the end of a learning sequence (Karst et al., 2017). In such educational settings, the term *assessment* typically refers to the process of collecting and interpreting information about learning-relevant student and task characteristics, which is then condensed into a *judgment* aimed at informing educational decisions (Brunswik, 1955; Herppich et al., 2018). Teachers who are able to successfully cope with situations that require such assessments have been found to ensure higher instructional quality (Blömeke et al., 2020; Helmke & Schrader, 1987) and higher learning gains for students (Brunner et al., 2013; Karst et al., 2014). Therefore, teachers' skills that are related to successfully coping with such assessment situations (e.g., by making accurate judgments), or so-called *assessment skills*, are considered important prerequisites for high teaching quality (Schrader, 1989). This is also reflected in educational theories on teacher competences (Baumert & Kunter, 2006) and manifested in legal requirements of teacher education (Council of Chief State School Officers, 2013; Kultusministerkonferenz, 2004), setting assessment skills as important learning goals in teacher education.

With regard to the question of how well teachers' assessment skills are actually developed, early empirical studies indicated a need for improvement with regard to teachers' grading practices (Starch & Elliott, 1912, 1913). Although research today considers a broader variety of assessment situations (Karst et al., 2017), it still indicates that "teacher judgments are far from perfect and […] there is plenty of room for improvement" (Südkamp et al., 2012, p. 755). As teachers in the early years of their professional careers are particularly challenged by the variety of tasks involved in teaching practice, including student assessment (Stokking et al., 2003), implementing learning environments that allow an *approximation of practice* in university teacher education seems promising (Grossman et al., 2009).

Computer-based simulations that allow learners to deliberately practice skills within simulated situations have been effectively used for the training of practical skills (Petko, 2014). One of the first examples of a computer-based simulation for fostering teachers' assessment skills is the simulated classroom (Südkamp et al., 2008). In this simulation, teachers have to assess the mathematical knowledge of 10 simulated third graders. To gather relevant information about the students, future teachers have 18 minutes to select questions (out of a set of 40 questions) and pick students to answer the selected question. Recently, several other simulations have been developed (Fischer & Opitz, 2022).

The emerging development of simulations raises the question of how learning assessment skills in such simulations happens and how it can be better supported. As one way to increase learning, this dissertation focuses on scaffolding as a supporting measure that can be added to many simulations as instructional support, even after they have been released for use (Chernikova, Heitzmann, Stadler, et al., 2020; Quintana et al., 2004). With scaffolding embedded in the simulation, learners may still work in the simulated situation on their own but are assisted by the simulation taking over certain subtasks (Wood et al., 1976). For example, a scaffold in the simulated classroom (Südkamp et al., 2008) could take over the selection of relevant questions so that future teachers can focus on picking students and interpreting their answers. One crucial feature of scaffolding is that it supports learners exactly where they need it (Collins et al., 1989). Therefore, possible directions of support can be manifested in stable learner characteristics (e.g., learners' prior knowledge), as well as in the learners' performance and situative learning experiences within the simulation, such as the learners' current motivation (Aleven et al., 2017). In each case, thorough considerations preceding the design and development of scaffolding are necessary to ensure that scaffolding is tailored to learners' needs. Ideally, such preceding considerations are not only based on theoretical frameworks but also include empirical studies (Plass & Pawar, 2020). Empirical studies may provide additional insights into how learners learn within a given simulation and possibly point to directions for support measures. To my knowledge, however, this approach of combining theoretical and empirical perspectives for developing effective scaffolding has rarely been taken in recent research on scaffolding in computer-based settings, perhaps due to a lack of resources (e.g., time, money, personnel) or the absence of a standardized methodology to develop effective scaffolding.

The present dissertation provides an example of how effective scaffolding in the context of teachers' assessment skills can be developed and may thus serve as a blueprint for other researchers and simulation designers. In that regard, this dissertation goes beyond the mere evaluation of more or less thoroughly designed scaffolds by capturing theoretical, empirical, and methodological perspectives for developing scaffolding that supports learners where they need it. Theoretical considerations serve as a starting point from which the learner characteristics of future teachers important for learning assessment skills in simulations are derived. These considerations are followed by empirical investigations of what individual learning prerequisites future teachers bring to the learning environment, which allows, in turn, to determine possible directions for scaffolding measures. Drawing on a second empirical study, the developed scaffolds are then evaluated with regard to their general effectiveness. Altogether, the present dissertation provides insights into scaffolding and its development to support (future) teachers in acquiring assessment skills in simulations. It also provides further insights into assessment skills and learning in simulations.

2 Theoretical Background

2.1 Teachers' Assessment Skills

According to Artelt and Gräsel (2009), assessment skills[1] refer to adequately assessing student characteristics and task requirements. In other words, a skilled teacher in terms of student assessment is one who assesses students adequately. This represents a *holistic* view defining good assessment skills by effective assessment performance (i.e., adequate assessment). This stresses that performance is the main criterion for the validity of measuring teachers' assessment skills. Further, as performance is linked to the given situation, this holistic view also reflects that assessment skills involve situative and thus domain- and topic-specific aspects (Spinath, 2005). Recently, an *analytic* perspective stemming from the notion of competence has been added to the discourse on teachers' assessment skills. From this perspective, teachers' assessment skills are considered dispositions to successfully deal with the process of assessing learning-relevant student characteristics (Herppich et al., 2018). This perspective bears the advantage that assessment skills are considered rather stable, learnable, and trainable. However, Blömeke, Gustafsson, and Shavelson (2015) highlighted that both the analytic and the holistic perspectives are necessary to validly describe and conceptualize teachers' assessment skills, and that both perspectives are linked via situation-specific skills. Thus, even though recent conceptualizations of assessment skills rather take an analytic perspective, the role of situation-specific skills and assessment performance for teachers' assessment skills is acknowledged, and both aspects are included in these conceptualizations (Heitzmann et al., 2019; Herppich et al., 2018; Loibl et al., 2020). Thus, recent conceptualizations of assessment skills include three aspects: dispositions, situation-specific skills, and assessment performance (see also Figure 1).

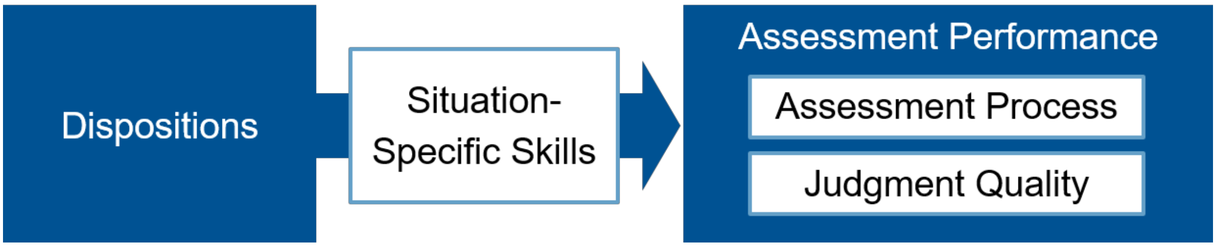

Figure 1. Conceptualization of assessment skills based on Blömeke, Gustafsson, and Shavelson (2015).

[1] Note that current research also uses other terminology to describe teachers' assessment skills, such as diagnostic skills, diagnostic competence, assessment competence, or assessment literacy, see Herppich et al. (2018).

With regard to the dispositions in these conceptualizations, it is often drawn mostly on cognitive dispositions, such as professional knowledge, but the role of dispositions, such as motivational characteristics, is acknowledged (Herppich et al., 2018; Loibl et al., 2020). However, cognitive and motivational dispositions are not only part of conceptualizations of assessment skills but can also be seen as learning prerequisites that influence learning in simulations (Heitzmann et al., 2019). These dispositions will be discussed in Section 2.3.1 in greater detail.

As another part of teachers' cognition, situation-specific skills, such as perception, interpretation, and decision-making, are expected to mediate between dispositions and assessment performance (Blömeke, Gustafsson, and Shavelson, 2015). The distinction between future teachers' situation-specific skills and their general dispositions has also been supported empirically (Blömeke et al., 2016) and included in current conceptualizations of assessment skills (Loibl et al., 2020).

With regard to assessment performance, prior research has often focused on the quality of the assessment. In their definition of teachers' assessment skills, Artelt and Gräsel (2009) required *adequate* assessments. However, the definition does not specify which criteria must be fulfilled for an assessment to be adequate, yet such criteria are essential for measuring assessment performance. Typically, assessments are required to be objective, reliable, and valid (Schrader, 2013). Whereas earlier research on assessment skills has focused on objectivity and reliability (Ingenkamp & Lissmann, 2008), present research draws mainly on criterion validity for determining the adequacy of an assessment (see Kane & Wools, 2020). Therefore, researchers often compare a teacher's judgment to an expert's judgment or to students' results from a standardized test (Karst et al., 2014). The result of this comparison is referred to as teachers' judgment accuracy, which can be considered the most prominent measure of judgment quality. Judgment accuracy can, for example, be expressed as the percentage with which a teacher and the criterion agree (Urhahne & Wijnia, 2021). Three other components of judgment accuracy were introduced by Schrader (1989): the level component, the rank component, and the differentiation component. The level component requires an assessment situation in which students need to be assigned to certain levels (e.g., skill levels) to analyze the extent and direction of disagreement between the teacher's judgment and a certain criterion. This allows for, for example, determining whether teachers overestimate or underestimate their students. The differentiation component analyzes whether teachers are able to correctly approximate the variance of the students' characteristics to be assessed. The rank component analyzes whether teachers are able to correctly rank their students according to the students' characteristics to be assessed.

All these measures allow researchers to quantify assessment performance, operationalized by judgment accuracy as the typical assessment quality indicator (Urhahne & Wijnia, 2021). However, to gain a better understanding of how teachers' come to certain judgments, in addition to considering teachers' dispositions and their situation-specific skills, it seems reasonable to investigate how teachers' dispositions come to action by the application of situation-specific skills (Blömeke, Gustafsson, & Shavelson, 2015). This becomes evident in the assessment process (Schrader, 2013; Wildgans-Lang et al., 2020). Therefore, current conceptualizations of assessment skills explicitly include and conceptualize the assessment process as a second aspect of assessment performance (Heitzmann et al., 2019; Herppich et al., 2018). These conceptualizations differ in their granularity. Herppich et al.'s (2018) model is more general and points out general constituents of the assessment process, such as the *processing of given information*. It also highlights two processing modes based on dual processing theories: the intuitive mode, in which information is automatically processed, and the reasoning mode, in which information is processed more systematically (Ferreira et al., 2006). The model of Heitzmann et al. (2019) is more fine-grained (regarding the processing of given information) by specifying eight epistemic activities based on Fischer et al.'s (2014) framework for scientific reasoning carried out during the assessment process. These activities have also been related to the concept of professional vision and noticing (Kramer, Förtsch, Seidel, & Neuhaus, 2021; Seidel & Stürmer, 2014; van Es & Sherin, 2002). However, so far, there has been no generally accepted way of conceptualizing and analyzing the assessment process. Further, approaches to the assessment process differ not only conceptually but also methodologically: some researchers draw on think-aloud protocols (e.g., Oudman et al., 2018) to gain insights into the assessment process, while others use log data during a computer-based simulation (e.g., Brandl et al., 2021), or eye-tracking (e.g., Kosel, Holzberger, & Seidel, 2021). Analyzing notes taken while assessing students has also been shown to be a valid way of approaching an understanding of the assessment process (e.g., Codreanu, 2021).

This broader understanding of student assessment, including dispositions, situation-specific skills, and their manifestation in the assessment process, can then lead to a further understanding of shortcomings in teachers' judgments. Based on this understanding, current learning opportunities may be improved, and new learning opportunities may be designed. This also appears to be necessary. Recently, Urhahne and Wijnia (2021) reviewed findings on teachers' judgment accuracy from 1980 to 2019. For example, regarding academic achievement, teachers tended to overestimate students' performance on a standardized test (level component), whereas no signs of systematic deviations regarding the differentiation component were found (Urhahne & Wijnia, 2021). Furthermore, the rank component (operationalized as the correlation between teachers' judgments and student academic achievement) has been reported to be rather high ($r = .63$) but still requires improvements (Südkamp et al., 2012). It is worth noting that these results do not consider teachers' professional experience. For example, research with a particular focus on teachers at the beginning of their professional careers has pointed toward an overload with regard to tasks related to student assessment (Levin et al., 2009).

2.2 Simulation-Based Learning

Future teachers may benefit from additional learning opportunities that can be provided in university teacher education, such as simulations (Blömeke, Hoth, et al., 2015; Chernikova, Heitzmann, Stadler, et al., 2020). In such educational contexts, a simulation typically refers to environments that reflect an aspect of (professional) practice, and have the feature of being interactive (Petko, 2014). That is, learners can interact with one or multiple agents within a simulation to mimic professional practice (Cook et al., 2013, p. 874). Such simulations can be used for teaching and training purposes. As a learning environment, simulations are fairly proximal to practice and may thus enable learners to more easily recognize patterns within a practice situation that they might have already dealt with in a simulation, thus allowing for the transfer of knowledge acquired in the simulation (Kahneman, 2003; Kolodner, 1992; Tversky & Kahneman, 1974). With regard to assessment skills, different simulations have been used thus far, ranging from live simulations (e.g., Fink et al., 2021; Kron et al., 2021) to computer-based simulations (e.g., Südkamp et al., 2008), also including video-based simulations (e.g., Codreanu et al., 2020).

2.2.1 Video-Based Simulations

As interactive learning environments that mimic aspects of real professional practice, simulations can be regarded as *approximations of practice*, which Grossman et al. (2009) defined as "opportunities for novices to engage in practices that are more or less proximal to the practices of a profession" (p. 2058). This definition highlights the existence of different degrees of approximation to practice, depending on the proximity of the learning environment to practice. In the past few years, researchers have started developing simulations in the sense of approximations of practice (Codreanu et al., 2020). In their call for approximations of practice, Grossman et al. (2009) also targeted current university teacher education, in which training situations for the acquisition of complex teaching skills, such as assessment skills, are scarce and mainly happen in the form of practicums at school (Cohen et al., 2013). However, such practical phases seem to provide particularly few learning opportunities for future teachers to engage in activities related to assessment skills (Weresch-Deperrois et al., 2009), and if they do, the active teaching experiences as a central learning opportunity within these practical phases (see Arnold et al., 2014) may be too overwhelming for future teachers (Dicke et al., 2016). Thus, additional learning opportunities within university teacher education that allow future teachers to focus exclusively (or at least primarily) on assessment skills seem promising. Grossman et al. (2009) referred to this kind of training as a decomposition of practice, which "involves breaking down practice into its constituent parts for the purposes of teaching and learning" (p. 2058). For example, in teaching practice, teachers typically have to simultaneously deal with student assessment and classroom management. A decomposition of practice, however, would allow future teachers to focus on one of these tasks at a time and to be more deliberate in their practice (Ericsson et al., 1993; Grossman et al., 2009).

Therefore, learning environments may be particularly beneficial for future teachers when they simultaneously represent approximations and decompositions of practice. Again, simulations can provide opportunities for future teachers to authentically approximate practice and target specific skills, such as assessment skills (Chernikova, Heitzmann, Stadler, et al., 2020; Codreanu et al., 2020). Among different types of simulations, video-based simulations may be particularly promising, as they combine the advantages of simulations (e.g., interactivity) with the advantages of the use of authentic videos: The use of classroom videos not only preserves authenticity in a scalable way (Brophy, 2004) but also allows for a decomposition of practice. For example, by assessing students in scripted videos that depict students working on a task without any non-task-related behavior, future teachers can facilitate their assessment skills without being distracted by other common requirements in classroom practice. Further, by varying the number of students to be assessed in these scripted videos, the task difficulty could be adjusted to the future teachers' zone of proximal development (Fischer et al., 2022; Vygotsky, 1978). Hence, it is not surprising that an increasing number of researchers are drawing on videos to facilitate teaching skills (Borko et al., 2008; Gold et al., 2013; Santagata, 2009), and are substantiating the premises of authentic videos with regard to learning gains (Gaudin & Chaliès, 2015; Stürmer et al., 2013).

To summarize, beyond being an always available and scalable learning environment, video-based simulations can have the advantage of being an approximation of practice while allowing for decomposing practice (Grossman et al., 2009; Stürmer et al., 2013). Both advantages, however, can also be seen as requirements for effective video-based simulations, as they influence the learners' *situative learning experiences*—that is, their perceptions and experiences in the video-based simulation (Chernikova, Heitzmann, Stadler, et al., 2020; Grossman et al., 2009).

2.2.2 Situative Learning Experiences in Video-Based Simulations

Situative learning experiences are considered important situative prerequisites of learning and include, among others, the learners' motivation and cognitive load while completing the video-based simulation, as well as the learners' subjective sense of presence and authenticity.

Regarding the analysis of learning processes, motivation is among the most prominent measures of situative learning experiences. Research on motivation and its role in learning has a long tradition in educational psychology (Bandura, 1977; Heckhausen, 1967). Within virtual learning environments, such as video-based simulations, motivation is expected to be an important predictor of performance and learning (Heitzmann et al., 2019). For example, Chen and Wu (2012) found aspects of learners' motivation to predict the use of metacognitive strategies ($\beta = 0.12$), which then influenced the learners' performance. According to the expectancy-value theory of motivation (Wigfield & Eccles, 2000), a learner's motivation for a given activity is influenced by two main components: the learner's expectancy to successfully accomplish the activity (*success expectancy*) and the subjective value the learner attaches to that given activity (*value*). This model also specifies that success expectancy is influenced by, among others, the learners' perceptions of the demands of the activity. For example, a higher teachers' confidence about student assessment may lead to an increased motivation for student assessment in the classroom. By contrast, the perception of the utility of the given activity (*utility value*) influences the teachers' subjective value component. For example, being aware of the relevance of student assessment for teaching practice may also lead to increased motivation for student assessment in the classroom (Wigfield & Eccles, 2000). Besides success expectancy and value, their interaction is also expected to play a role in learners' motivation

(Nagengast et al., 2011; Wigfield & Eccles, 2000). In Trautwein et al.'s (2012) study with 2508 secondary school students, comparably high scores on both components led to the highest scores on a mathematics test, whereas comparably low scores on the success expectancy component or the value component could not be compensated by comparably high scores of the other component. Interestingly, the combination of comparably high subjective value and comparably low success expectancy led to even worse mathematics test scores than comparably low scores on both components. However, other interaction patterns have also been found (e.g., Song & Chung, 2020). Against this backdrop, the role of the interaction between success expectancy and value remains an open question (Wigfield & Eccles, 2020).

Besides motivational situative learning experiences, cognitive situative learning experiences are also relevant for learning in general, and for learning in video-based simulations in particular (Mayer, 2014). For learning with virtual environments, researchers typically take the construct of cognitive load into account (Sweller et al., 1998; Sweller et al., 2019). Cognitive load theory draws on the premise that humans' working memory has a limited capacity. During the learning process, these limited working memory resources can be distributed between intrinsic cognitive load, extraneous cognitive load, and germane resources (Sweller et al., 2011). Intrinsic cognitive load is imposed by the complexity of the learning material itself. In the context of a video-based simulation for teachers' assessment skills, intrinsic cognitive load refers to the working memory load, which is imposed due to the complexity of student assessment for the teacher (e.g., information that is processed about a student). Extraneous cognitive load refers to the space in the working memory load that is imposed due to the presentation of the learning material. For example, navigating through a complex and unstructured video-based simulation may impose a high extraneous cognitive load (Sweller et al., 1998). Based on the findings of previous studies, Sweller et al. (2011) changed the initial term germane cognitive load to *germane resources*, which are now understood as "working memory resources that are devoted to information that is relevant or germane to learning" (p. 57). Such information imposes an intrinsic cognitive load, which underpins that intrinsic cognitive load and germane resources are intertwined and necessary for learning. Thus, intrinsic cognitive load needs to be kept at an appropriate level, and extraneous cognitive load needs to be kept as low as possible to have enough germane resources for learning. This means that, for example, a video-based simulation's user interface should be designed learner-friendly, and the complexity of the given learning material should be adjusted to an appropriate level that provides enough learning content but is not overwhelming for the learners (Grossman et al., 2009; Sweller et al., 2019).

Whereas motivation and cognitive load are typical measures for situative learning experiences in general, the feeling of presence and the perception of authenticity become particularly important in virtual environments, such as video-based simulations. Schubert et al. (2001) described presence as the “sense of being in the virtual environment” (p. 266). Sometimes, presence is also understood as mental or cognitive involvement (Stevens & Kincaid, 2015; Witmer & Singer, 1998). The sense of presence is subjective and influenced by the design of the video-based simulation. For example, a video-based simulation that is very proximal to practice can enlarge the sense of being in the simulated situation of practice (Grossman et al., 2009; Schubert et al., 2001). Evidence that learners‘ perception of presence facilitates performance in simulations may not be surprising (Selzer et al., 2019; Stevens & Kincaid, 2015). For example, Stevens and Kincaid (2015) found a moderate relation ($r = .22$) between the sense of presence and performance for novice soldiers in a military simulation. The concept of presence is also linked to the perception of authenticity because the high authenticity of a video-based simulation is expected to positively affect the sense of presence in this simulation (Gilbert, 2016). Authenticity refers to the physical resemblance of the simulated situation with the real-world situation but also to the functional correspondence with regard to the simulated task (Chernikova et al., in press). Authenticity has also been linked to learning. In their meta-analysis, Chernikova, Heitzmann, Stadler, et al. (2020) found simulations with low authenticity ($g = 0.58$) to be less effective for learning than simulations employing higher levels of authenticity ($g = 0.86$). Thus, besides motivational and cognitive situative learning experiences, presence, and authenticity are important aspects of learners’ perceptions and experiences when learning in a video-based simulation.

2.3 Facilitating Teachers’ Assessment Skills in Simulations

Since teachers’ assessment skills are relevant to multiple aspects of teaching that have been identified as core practices of teaching (Grossman, 2018), facilitating teachers’ assessment skills in university teacher education may contribute to high teaching quality in future teachers’ later professional practice. To this end, video-based simulations as approximations and decompositions of practice are considered promising for facilitating assessment skills (Grossman et al., 2009). In this regard, several different simulations have been developed

(Fischer & Opitz, 2022; Südkamp et al., 2008). To guide research on facilitating assessment skills in simulations, the COSIMA research unit (2021) developed a framework that outlines the components of learning assessment skills in simulations (see Figure 2).

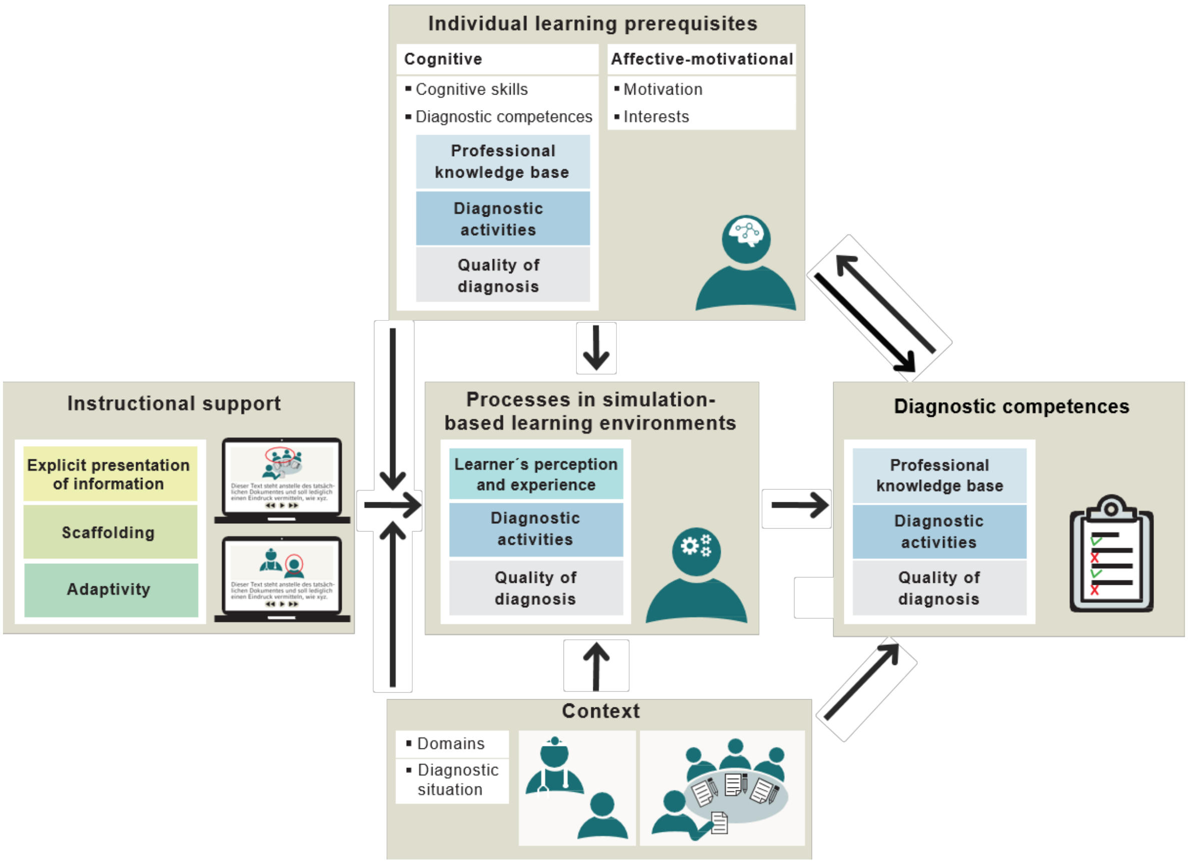

Figure 2. COSIMA framework[2] *(see also Heitzmann et al., 2019).*

This framework comprises five main components: First, and most importantly, the learners' *assessment skills* (labeled diagnostic competences in this framework) as an outcome of the simulation. For example, this dissertation specifically focuses on assessment performance as a learning process/outcome measure as described in Section 2.1. Second, the *processes in the simulation-based learning environment* comprise the diagnostic activities and the quality of the assessment in the simulation, as well as the learners' situative learning experiences. Note that the term diagnostic activities in the framework refers to the conceptualization of goal-oriented activities conducted during the assessment process described in Section 2.1 (Fischer et al., 2014; Heitzmann et al., 2019). To increase learning gains, the processes in the simulation can be promoted by additional *instructional support,* the third component of the framework. The framework specifies, among others, scaffolding as a means of instructional support. Fourth, the

[2] From *CC BY licenses for the COSIMA framework* by COSIMA research unit, 2021 (https://www.en.for2385.uni-muenchen.de/cosima_framemodel1/ccbylicenses/index.html). CC BY 4.0.

framework takes the learners' *individual learning prerequisites* into account. It considers motivational-affective as well as cognitive prerequisites. Future teachers' assessment skills before the simulation are also considered individual learning prerequisites. These learning prerequisites are expected to influence teachers' future assessment skills, the processes in the simulation, and the effects of instructional support for the processes in the simulation. To distinguish between prerequisites that appear consistently in different learning environments and prerequisites that may vary from one learning environment to another, I refer to rather stable prerequisites as learner characteristics (e.g., dispositions, such as interest or professional knowledge) and consider other prerequisites, such as motivation, as situative learning experiences. Lastly, the framework highlights that assessment skills, and thus simulations for facilitating assessment skills, are influenced by context variables, such as the domain of the assessment.

In their meta-analysis, Chernikova, Heitzmann, Stadler, et al. (2020) investigated the effectiveness of such simulations and found simulations to be an effective means of fostering complex skills, such as teachers' assessment skills, with a mean effect of $g = 0.58$. Further, the effectiveness of such simulations varied along the dimensions of *learner characteristics* (e.g., when using familiarity with the context as an indicator of prior knowledge, they obtained effects of $g = 0.83$ for learners with high prior knowledge and $g = 0.67$ for learners with low prior knowledge) and the employed types of *scaffolding* (e.g., the combination of prompts and worked-out examples reached an effect of $g = 1.60$, whereas simulations without scaffolding had an effect of $g = 0.88$). Both aspects, learner characteristics and scaffolding, are discussed in the following two subsections.

2.3.1 Learner Characteristics

Regarding learner characteristics, researchers often draw on cognitive and motivational variables as important prerequisites for effective learning (Kosel, Wolter, & Seidel, 2021). Even though theoretical frameworks may also include executive functions (Heitzmann et al., 2019), a general information processing ability (Herppich et al., 2018) or more general cognitive abilities (Loibl et al., 2020), teachers' professional knowledge is highlighted by several frameworks and is typically regarded as a main component of cognitive learner characteristics. Shulman's (1986) framework provides guidance to structure teachers‘ professional knowledge in general, which is also a relevant structure of knowledge underlying teachers‘ assessment performance (Förtsch et al., 2018): Teachers generally require content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical-psychological knowledge (PK). CK refers to teachers' knowledge of the subject domain itself, and PCK refers to knowledge of learning and teaching within a subject domain. Thus far, CK and PCK are considered domain-specific (Förtsch et al., 2018) and may thus be particularly important for assessment performance in strongly content-related settings (Hoth et al., 2016). For example, when assessing students' performance on a geometry task (as a highly content-related setting), it seems particularly important that teachers are able to solve the task by themselves (referring to CK) and that teachers know about typical student misconceptions (referring to PCK). By contrast, PK, defined as domain-general knowledge referring to general pedagogical-psychological aspects in school classrooms, such as classroom management or judgment biases, may be less relevant in this situation. However, PK may be particularly important for assessment performance in general instructional settings and teacher actions.

Empirically, not all prior studies find significant positive effects of all professional knowledge facets (CK, PCK, PK, and their possible subfacets) for high judgment quality[3], yet the tendency of empirical results substantiates the role of these facets depending on the domain specificity of the assessment situation. Studies with a general focus on teaching (e.g., teachers' assessment of the instructional quality of a lesson) have reported significant effects of teachers' PK on their assessment performance (König et al., 2014; Kramer, Förtsch, Boone, et al., 2021). Studies with a stronger content-related focus (e.g., teachers assess the learning potential of a mathematical task) did not find significant main effects of PK but a significant main effect of teachers' domain-specific CK on teachers' assessment performance (Aschbacher & Alonzo, 2006; Kron et al., 2022; Todorova et al., 2017). With regard to teachers' PCK, most studies report significant effects on assessment performance, ranging from weak influence (McElvany et al., 2009) and medium (Rieu et al., 2022) up to larger effects of PCK with regard to specific aspects of judgment quality (Ostermann et al., 2018).

In summary, prior research offers tendencies toward positive effects of professional knowledge on the quality of teachers' judgments. With regard to teachers' motivational learner characteristics, the number of conclusive empirical studies is small, and there is no clear tendency regarding what role such characteristics can play in teachers' assessment performance. However, as motivational teacher characteristics, such as teacher‘s interest and self-efficacy, have been found to be important for effective teaching in general (Schiefele et al.,

[3] See, for example, Karing, 2009, for knowledge in general or McElvany et al., 2009, for different PCK subfacets.

2013; Tschannen-Moran et al., 1998), they are also expected to play a role in teachers' assessment performance (Heitzmann et al., 2019; Herppich et al., 2018; Klug et al., 2016; Loibl et al., 2020). In one of the few conclusive empirical studies, Kron et al. (2022) pointed to teachers' interest as a *door-opener* for the application of professional knowledge. In their study, they found a significant interaction between PCK and interest in judgment accuracy; future teachers with a higher PCK and higher interest also reached a higher judgment accuracy. Further, contrary to their theoretical expectations, Klug et al. (2016) found a negative effect of future teachers' self-efficacy on their assessment performance, which they explained with the negative effects of overestimation. Thus, both interest and self-efficacy could play a role in teachers' assessment performance, which needs to be further investigated. Besides their influence on teachers' assessment performance, these two variables can also be seen as important prerequisites for learning assessment skills (Ainley et al., 2002; Dupeyrat et al., 2011; Praetorius et al., 2013; Praetorius et al., 2016). Lastly, self-regulation[4] is often seen as an important prerequisite for learning in general (Dent & Koenka, 2016). Even though self-regulation has been shown to be an important teacher characteristic for effective teaching, its role has so far never been successfully investigated in the context of teachers' assessment skills (Klusmann et al., 2008).

To summarize, the theoretical assumptions that teachers' CK, PCK, and PK, as well as teachers' interest, self-efficacy, and self-regulation, can play an important role in teachers' assessment performance seem to be substantiated by empirical findings. Thus, particular knowledge about how these learner characteristics interact with performance and learning outcomes in video-based simulations can also serve as a basis for the development of supporting measures, such as scaffolding.

2.3.2 Scaffolding

Scaffolding is considered an effective means of supporting performance and learning in various domains (Belland et al., 2017; Hardy et al., 2019). Since its introduction by Wood et al. (1976), scaffolding theory has been further developed conceptually and has also been integrated into well-established theories of learning, such as cognitive apprenticeship (Collins et al., 1989). The value of scaffolding is also underlined by empirical findings that positively evaluate the effectiveness of scaffolding in a large number of learning settings (Belland et al., 2017; van de

[4] This dissertation focuses on motivational aspects of self-regulation; see Boekaerts (1999) for other components. For the sake of readability, I will continue to use the term self-regulation but refer only to its motivational aspects.

Pol et al., 2010). With regard to the use of video-based simulations in teacher education, scaffolding is considered promising for increasing performance and learning (Chernikova, Heitzmann, Fink, et al., 2020; Fischer et al., 2022).

The initial concept of scaffolding was introduced by Wood et al. (1976) in the context of tutor support for acquiring problem-solving skills. By the term *scaffolding process*, the authors referred to a tutor action "that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts" (p. 90). Therefore, the support of the tutor may, for example, consist of motivating the learners to solve a given task, or may consist of taking over subtasks that are either not feasible for the learners or occupy too much of the learners' capacities for carrying out the given task. These ideas refer to *recruitment* and a *reduction in degrees of freedom*, which have been pointed out by Wood et al. (1976) along with four other scaffolding strategies. Theoretical considerations on the concept of scaffolding have emerged since its introduction by Wood et al. (1976), moving from the description of scaffolding strategies toward general categorizations of scaffolding. For example, Hannafin et al. (1999) distinguished between four categories of scaffolding, such as conceptual scaffolding that aims at nudging learners to relevant considerations (e.g., providing additional knowledge).

Technological advances have inspired researchers to investigate scaffolding not only in terms of tutor actions but also in terms of actions in virtual learning environments, such as video-based simulations (Heitzmann et al., 2019; Quintana et al., 2004). Similar to real-life tutors, video-based simulations can take over tasks that are barely feasible or too demanding for learners (Belland et al., 2017). As a crucial feature of scaffolding in such virtual learning environments, the learners' responsibility for solving the given task themselves is highlighted. Even though learners are supported by scaffolding, they still have to complete the task themselves; they remain actively involved in and responsible for the final solution of the task (Belland, 2011). A prominent example of scaffolding in virtual learning environments that allows learners to remain responsible for solving a task is conceptual prompts.

2.3.2.1 Conceptual Prompts

In its conceptual advancement, scaffolding has also been related to Vygotsky's (1978) concept of zone of proximal development (Stone, 1998). For example, Hannafin et al. (1999) used the term "Vygotskian scaffolding" to refer to certain scaffolding strategies (p. 131), and Palincsar (1998) highlighted that scaffolding should fit the students' zone of proximal development. This relationship to the zone of proximal development, which often has a more cognitive connotation, may explain why scaffolding is often seen as cognitive support, even though half of Wood et al.'s (1976) original scaffolding strategies refer to motivational support (Belland et al., 2013; Hannafin et al., 1999). Thus, empirical investigations of scaffolding strategies mainly refer to cognitive support. Based on these empirical studies, prompts are considered to be among the most promising scaffolding strategies, particularly in virtual learning environments (Chernikova, Heitzmann, Stadler, et al., 2020).

Prompts can be defined "as recall and/or performance aids, which vary from general questions […] to explicit execution instructions" (Bannert, 2009, p. 139). Such prompts are also often used in virtual learning environments, as they have been reported to be effective and easy to implement (Bannert & Reimann, 2012), leading to a broad variety of different prompts (Berthold et al., 2007; Zumbach et al., 2020). Following the categorization of scaffolding by Hannafin et al. (1999), *conceptual prompts* represent one category of prompts[5], which can be understood as prompts guiding learners' considerations within a learning environment. The effectiveness of such prompts can also be related to cognitive load theory (Klepsch & Seufert, 2021), as conceptual prompts provide guidance to deal with the given task, which otherwise would have imposed cognitive load on the learners' working memory. The following example illustrates a possible connection of conceptual prompts to cognitive load: A prompt defining specific segmenting subtasks that were not given in the initial instruction can guide learners' considerations toward these subtasks and thus can be considered a conceptual prompt. This conceptual prompt changes the affordances of a given task (several subtasks instead of one extensive task), which may impose a lower intrinsic cognitive load. Conversely, such prompts also represent an additional element in the learning environment and thus lead to higher extraneous cognitive load (Mayer & Moreno, 2010). Depending on the available germane resources that are employed, these prompts can then lead to higher learning gains (Moreno & Mayer, 2010); see also Figure 3.

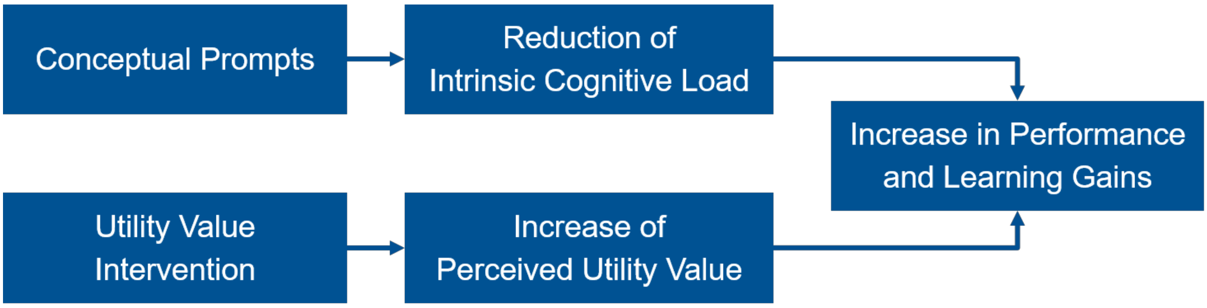

Figure 3. Possible effect mechanism of conceptual prompts and a utility value intervention.

[5] This category of prompts is sometimes called *cognitive prompts* (see Bannert, 2009; Wild and Schiefele, 1994).

Prompts were also used as effective means in the context of teachers' assessment skills (see Estapa & Amador, 2023 for an overview). For example, Irmer et al. (2022) found context-specific conceptual prompts activating future teachers' PCK to be effective regarding the scaffolded assessment performance and learning gains, whereas generic conceptual prompts activating more general strategies for the assessment process were not effective. Sommerhoff et al. (2023) compared the effects of two types of context-specific conceptual prompts in a video-based simulation. The first type drew future teachers' attention to a particular relevant event in a video and asked them to draw conclusions from this event about a corresponding facet of what was being assessed. This specific type of prompt was found to be generally effective for judgment following the scaffolded assessment process and with regard to learning gains (β = 0.78). The other type of prompt asked future teachers to make inferences about how two different facets of what is being assessed interact (without naming any specific video event). This interconnecting type of prompt was not found to be generally effective, neither for the judgment following the scaffolded assessment process, nor for a later judgment following a non-scaffolded assessment process; thus, no learning gains were indicated through these interconnecting prompts. However, in a further analysis, Sommerhoff et al. (2023) found that future teachers required sufficient PCK to effectively use such prompts. This seems reasonable, as the given interconnecting prompts can only support future teachers (and may impose additional cognitive load otherwise) once they have identified events in the video as relevant and connected them to the corresponding facet of what is being assessed, both requiring sufficient PCK (Meschede et al., 2017; Seidel & Stürmer, 2014).

In other words, future teachers' PCK turned out to be a moderator for the effectiveness of the given interconnecting prompts, and thus highlights what was also expected theoretically (Heitzmann et al., 2019): scaffolding can be differentially effective. The phenomenon of the differential effectiveness of prompts has also been found empirically in other domains, drawing on cognitive moderators, such as verbal intelligence, reading competence, or general domain knowledge (Horz et al., 2009; Pieger & Bannert, 2018). Motivational variables may also moderate the effectiveness of prompts. However, empirical studies investigating the possible moderating role of motivational variables are scarce within the research on prompts, even though the motivational effects of conceptual prompts seem plausible (Belland et al., 2013; Bixler, 2007). In motivational scaffolding research, considering motivational variables is quite common.

2.3.2.2 Utility Value Intervention

Compared to research on conceptual scaffolding, research on motivational scaffolding is underrepresented, despite the presence of promising scaffolding strategies (Belland et al., 2013; Lazowski & Hulleman, 2016). One promising example of motivational scaffolding—generally but also within video-based simulations—is *utility value interventions* (Hulleman & Harackiewicz, 2021). A utility value intervention is a motivational intervention that aims to increase the learners' utility value component of motivation by supporting them in connecting the learning content to their later professional lives. A study by Hulleman et al. (2010) is considered one of the first experimental studies to examine the role of interventions aiming at increasing utility value (Hulleman & Harackiewicz, 2021). In this study, psychology undergraduate students reflected on the utility of a math technique in a writing task, which led to a higher perceived utility value and interest in the math technique. Performance did not improve significantly in the study, but later studies also found effects on performance and further outcome measures (Durik et al., 2015; Harackiewicz et al., 2016). The utility value intervention in Hulleman et al.'s (2010) study draws on the *self-generation* strategy, which typically involves learners writing about a topic's connection to their own lives. This s*elf-generation* strategy can also be combined with another strategy—the direct *communication* of utility value (Gaspard et al., 2015)—which refers to presenting the relevance of a given activity directly to the learners. Empirically, the combination of both strategies, *self-generation* and *communication,* has yielded positive effects on interest, performance, and learning gains (Canning & Harackiewicz, 2015; Gaspard et al., 2021). By increasing the learners' motivation, particularly their utility value regarding the learning content (Wigfield & Eccles, 2000), utility value interventions may induce psychological processes, such as cognitive involvement, and development of interest (Hidi & Renninger, 2006), and behavioral processes, such as a higher engagement in the (learning) activity and improved performance. Both mechanisms may also influence each other, and may increase outcomes, such as grades, in the long term (Hulleman & Harackiewicz, 2021); see also Figure 3.

Concerning the analysis of the effectiveness of utility value interventions, researchers usually consider their differential effectiveness. For example, in Rosenzweig et al.'s (2020) study, a utility value intervention only had a small effect on exam scores (d = .30) but not for students with lower initial course performance: their post-intervention exam scores were significantly higher than for the control condition, and this difference was characterized by a large effect size of $d = .90$. Among possible moderators for the effects of utility value interventions, success expectancy, sometimes called confidence (Canning & Harackiewicz, 2015), or performance expectations (Hulleman et al., 2010) has been used very often (Gaspard et al., 2021). For example, with regard to utility value interventions using a *communication* strategy, learners with a comparably high success expectancy tended to benefit much more than learners with a comparably low success expectancy, who might not even benefit at all (Canning & Harackiewicz, 2015; Durik et al., 2015). Again, combining the *communication* strategy with a *self-generation* strategy allows for remediation, increasing the effects of the utility value intervention for learners with comparably low success expectancy (Canning & Harackiewicz, 2015).

With regard to teacher education, research investigating the effects of utility value interventions is scarce. This is surprising, as it is known that future teachers often perceive a lack of relevance regarding university learning content, particularly learning content with regard to educational sciences, for their later professional practice (Alles et al., 2019; Voss, 2022). In a recent article, Rochnia and Gräsel (2022) found future teachers' perception of the utility of educational science for their later professional lives to be medium (4.39 on a 7-point Likert scale). In this study, the authors employed a utility value intervention to increase this perception of utility. However, they did not find significant differences between the experimental and control groups regarding the perceived utility value after the intervention, which they also explained by the lack of concreteness of their intervention. It remains unclear whether utility value interventions can be used as effective interventions in teacher education, and for facilitating teachers' assessment skills, in particular.

3 The Present Research

Video-based simulations can be an effective and efficient way to facilitate teachers' assessment skills (Gaudin & Chaliès, 2015). Besides being fairly easy to implement in video-based simulations, Section 2.3.2 also highlights that scaffolding can be a very effective means for further optimizing performance and increasing learning gains in video-based simulations (Belland et al., 2017). However, Sections 2.3.2 also underpins that scaffolding is not necessarily effective and that effectiveness also depends on the learners' prerequisites that may be required to use scaffolding appropriately (Sommerhoff et al., 2023). Thus, for *effective* scaffolding, scaffolds need to match the prerequisites of the targeted learners. For example, scaffolding that supports learners in activating their knowledge is only *effective* if learners are unable to activate knowledge; otherwise, this kind of scaffolding is barely helpful and may even have negative effects (overscripting effects). Accordingly, for *effective* scaffolding, the following questions arise (Brown et al., 1982; Friedrich & Mandl, 1992; Plass & Pawar, 2020):

(1) What levels of relevant learner characteristics do future teachers bring to a video-based simulation designed to foster teachers' assessment skills, and how do these influence the processing of the simulation?
(2) What scaffolding strategies are effective for future teachers' with different learner characteristics?

These two questions are the superordinate research questions of this dissertation and are explicated in further detail in the following sections. This dissertation addresses both questions in two empirical studies, drawing on an already validated video-based simulation.

3.1 Future Teachers' Learner Characteristics

Article A provides an answer to the first question, which comprises three particular steps. As a first step, *relevant* learner characteristics need to be identified. Given the number of possible learner characteristics, a smaller set of variables needs to be selected. The set should contain only variables that show a considerable connection to teachers' assessment performance, as variables unrelated to assessment skills barely provide further insights with regard to the design of effective scaffolding. In that regard, the theoretical considerations of Heitzmann et al. (2019) and prior research (see Section 2.3.1) provide guidance: As the most important representative for cognitive learner characteristics, drawing on professional knowledge seems promising. Due to the context specificity of the assessment in the given video-based simulation, drawing on the teachers' CK and PCK as indicators of professional knowledge seems particularly promising. Regarding motivational characteristics, self-efficacy, interest, and motivational aspects of self-regulation are considered important learner characteristics that may also be relevant for assessment performance. Article A follows recommendations to complement this theory-based selection of learner characteristics with an empirical evaluation of their relevance for assessment performance and learning gains in the simulation (e.g., Plass & Pawar, 2020).

As a second step in answering question (1), it needs to be clarified to what extent future teachers bring the selected learner characteristics to the video-based simulation. As such learner characteristics can appear in complex patterns (Kosel, Wolter, & Seidel, 2021), a person-centered approach seems appropriate, which generally refers to a "clustering of people as opposed to variables" (Woo et al., 2018, p. 816). This is particularly promising for a cluster-wise development of effective scaffolding matching clustered learners' prerequisites (Tetzlaff et al., 2021). One such person-centered approach is latent profile analysis (LPA), which aims at identifying distinct learner "groups" that have similar scores across various indicators that are specified beforehand, referred to as "profiles" (Lazarsfeld & Henry, 1968). In the case of this dissertation, these profiles allow for a more detailed view of typical patterns in which different learner characteristics occur.

As a third step in answering question (1), these profiles need to be compared with regard to the processing of the simulation (here: situative learning experiences, assessment performance). This allows for the investigation of connections between learner characteristic profiles and teachers' assessment skills, but it also allows for the development of *effective scaffolding.* For example, based on which of these learner characteristic profiles positively affects judgment accuracy (or other aspects of the processing of the simulation), role-model profiles and important features of these profiles may be identified. These features may then be targeted by effective scaffolding. In summary, these considerations are reflected in the following research questions, which are addressed in Article A.

1.1. Do future teachers' relevant learner characteristics show distinct learner profiles? If so, which learner characteristic profiles can be identified for future teachers?
1.2. Do learner characteristic profiles affect future teachers' situative learning experiences in the simulation?

1.3. Do learner characteristic profiles affect future teachers' assessment process (watching videos, taking notes) and their final judgment (judgment report, judgment accuracy)?

With regard to these questions, the following was hypothesized in Article A: Given that future teachers' motivational learner characteristics correlate with each other (Holzberger et al., 2021), and different aspects of their professional knowledge also correlate (Riese & Reinhold, 2009), four profiles were expected within the explorative approach of LPA: two profiles with above-average (below-average) motivational learner characteristics and above-average (below-average) professional knowledge, and two profiles with above-average (below-average) motivational learner characteristics and below-average (above-average) professional knowledge. Among these, profiles with above-average motivational learner characteristics were expected to have higher motivational states in the video-based simulation and could perceive the video-based simulation as more authentic (Betz et al., 2016; Wigfield & Eccles, 2000). Profiles with above-average professional knowledge were expected to have more appropriate levels of cognitive load and could perceive the video-based simulation as more authentic (Betz et al., 2016; Sweller et al., 1998). Regarding the quality of the assessment process and the final judgment, motivational learner characteristics and knowledge may be equally advantageous for performance and learning in the simulation, and may compensate each other (if one is below average and the other above average) or be synergistic (if both are above average).

3.2 Effective Scaffolding Strategies

The answer to Research Question 2 is also based on an answer to Research Question 1. Having identified role-model profiles and important features of these profiles from the answer to Research Question 1 allows to derive possible effective scaffolding strategies that address those features. However, knowing scaffolding strategies that address these features does not necessarily lead to effective scaffolding; the scaffolding strategy itself must be effective in the given domain. The effectiveness of scaffolding is typically examined by the changes it induces in an outcome measure, such as teachers' judgment accuracy. Situative learning experiences may also be taken into account to explain these changes.

In his original definition of scaffolding, Wood et al. (1976) drew on two categories of scaffolds: conceptual and motivational scaffolds[6] (Belland et al., 2013). These two categories of scaffolds

[6] Nowadays, there are also other aspects of scaffolding, see Hannafin et al. (1999). However, considering all of these aspects is beyond the scope of article B.

are capable of addressing the key learner characteristics in the COSIMA framework (COSIMA research unit, 2021). Conceptual prompts as a representative of conceptual scaffolds and a utility value intervention as a representative of motivational scaffolds may be promising with regard to teachers' assessment skills (Hulleman & Harackiewicz, 2021; Sommerhoff et al., 2023). Since research on utility value interventions highlights the role of learners' success expectancy as a moderator of the effectiveness of such interventions, it seems important to also investigate whether this differential effectiveness transfers to the domain of teachers' assessment skills. This is particularly interesting regarding prompts, as the role of success expectancy in the effectiveness of conceptual prompts has rarely been investigated. Thus, Article B draws on conceptual prompts and a utility value intervention as scaffolding strategies for facilitating teachers' assessment skills (operationalized as judgment accuracy). Against this backdrop, Article B investigated the following questions:

2.1. Can conceptual prompts and/or a utility value intervention promote intrinsic cognitive load, extraneous cognitive load, utility value, and judgment accuracy?
2.2. Does learners' comparably high or comparably low initial success expectancy influence the scaffolds' effectiveness on judgment accuracy? Are the scaffolding conditions more effective in supporting the judgment accuracy of learners with comparably high or comparably low success expectancy?

With regard to these questions, the following was hypothesized in Article B: Regarding Research Question 2.1, it was expected that conceptual prompts reduce intrinsic cognitive load and lead to increased judgment accuracy. It was also expected that extraneous cognitive load would increase slightly, as conceptual prompts may be perceived as an additional element of the simulation that needs to be processed. The utility value intervention, in turn, was expected to increase learners' perceptions of utility value regarding assessment skills and increase judgment accuracy. The combination of both scaffolds was expected to be especially helpful by combining the advantages of both individual scaffolds. Regarding Research Question 2.2, conceptual prompts were expected to particularly boost the success expectancy of learners who lack success expectancy before the simulation. Learners who were confident in completing the assessment task were expected to benefit from a utility value intervention. The concrete design of both scaffolds is discussed in the following section, which describes the methodology.

4 Methodology

Both overarching research questions were investigated using a video-based simulation. This simulation is described along with the included measures and instruments, the design of the studies, the data collection procedure, and the data analysis in the following section.

4.1 The Visit-Math Project

This dissertation originates from the *Visit-Math* project, which is funded by a grant from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) to Tina Seidel, Daniel Sommerhoff, Stefan Ufer, Birgit Neuhaus, and Ralf Schmidmaier (grant numbers 1397/11-1 and 1397/11-2, FOR 2385). This project is embedded in the research group COSIMA, which aims to facilitate assessment skills in teacher and medical education by using simulation-based learning environments. In COSIMA, researchers from different domains (medical education, mathematics education, biology education, physics education, and educational psychology) collaborate to design and evaluate simulation-based learning environments that facilitate future teachers' and future physicians' assessment skills. The COSIMA research group includes seven subprojects, all of which have developed simulations (Fink et al., 2021; Kramer et al., 2020; Pickal et al., 2022; Radkowitsch et al., 2020; Stürmer et al., 2021; Wildgans-Lang et al., 2020), and one subproject to synthesize findings from the other subprojects (Fischer & Opitz, 2022). The simulation that was developed in the *Visit-Math* project is video-based and focuses on a typical assessment situation within a mathematical context (Codreanu et al., 2020; Codreanu et al., 2022). With regard to the goals, COSIMA aims to contribute to the validation of current conceptualizations of assessment skills and facilitating assessment skills in simulations, as indicated by the framework in Section 2.3 (COSIMA research unit, 2021; Heitzmann et al., 2019). Among other goals, COSIMA aims to develop and evaluate different types of instructional support, as has already been reported within *Visit-Math* (Sommerhoff et al., 2023) and in other projects within COSIMA (Irmer et al., 2022; Schons et al., 2022). As a further goal, establishing adaptive support based on the assessment performance is on the table, as reported by an article of Radkowitsch et al. (2021) that investigates the role of adaptive support. Altogether, this research allows for a further understanding of how to conceptualize and how to facilitate assessment skills, to which this dissertation also contributes.

4.2 Video-Based Simulation

Within the *Visit-Math* project and prior to this dissertation, a simulation had already been developed and validated as an approximation of practice (Codreanu, 2021). In this simulation, the participants are asked to make a judgment about two students' mathematical proof skills. Due to the scalability and positive effects of video-based formats for learning practical skills in teacher education (Seidel et al., 2011; Seidel et al., 2013), the project team chose the simulation to be video-based using scripted videos. These videos form the foundation of the assessment process in the video-based simulation. Besides the assessment process, the simulation also entails a *familiarization* part at the beginning, and the *final judgment* part at the end.

During *familiarization*, participants are familiarized with the fictional context and requirements of the simulation. They are asked to imagine themselves as interns at school observing a mathematics lesson, which represents a typical situation that future teachers face during university teacher education (Cohen et al., 2013). In the observed lesson, the simulated students work on one geometrical proof task on their own (see Figure 4), being supported by the teacher in some parts; again, this represents a typical situation in mathematics classrooms (Leiss, 2010). Further, during the familiarization part, participants are provided with the students' proof task and its solution. Participants are also provided with the assessment task (i.e., to assess the mathematical proof skills of two students) as a written text and as a video to technically familiarize them with the simulation. Participants are provided with descriptions of relevant indicators of students' mathematical proof skills. As indicators, mathematical content knowledge, methodological knowledge, and problem-solving strategies are used, as they have been found to be relevant for students to conduct mathematical proofs (Sommerhoff, 2017).

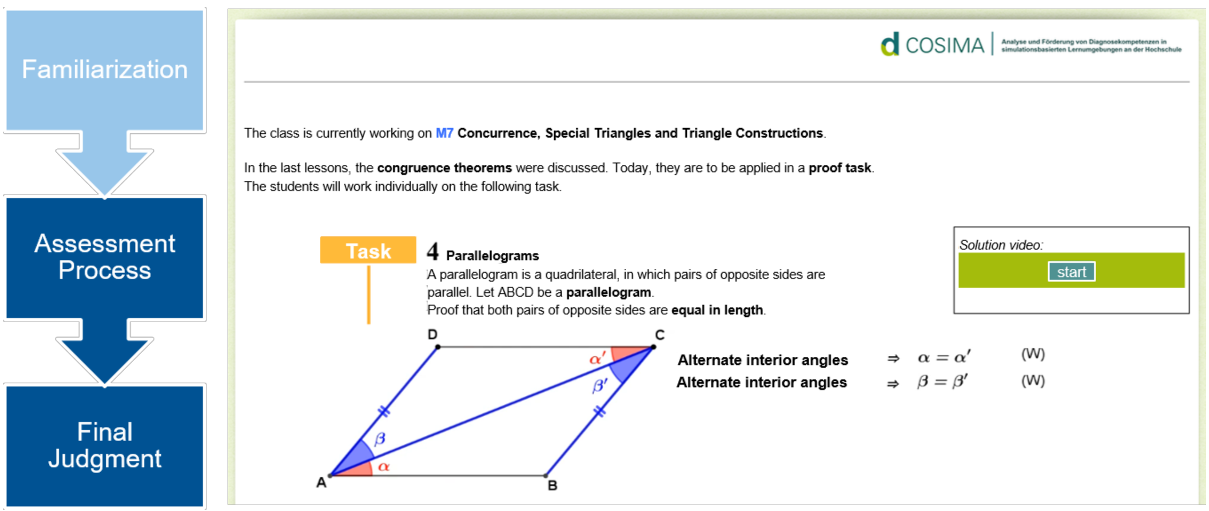

Figure 4. Presentation of the student task and its solution during familiarization.

During the *assessment process*, participants observe the two simulated students[7] in short, one-minute videos that depict a single simulated student who discusses his/her progress on the mathematical proof task with the teacher (see Figure 5) as a typical aspect of classroom work phases (Leiss, 2010). The videos alternate in the depicted simulated student: After a video with one simulated student, a video of the other simulated student is shown. The videos can be watched only once and without any pause, as is also the case in real classroom situations. Once participants think that they have gathered enough information about the simulated student, they can decide to stop watching videos with this student and continue watching videos of the other simulated student—representing a typical classroom situation in which a teacher has to decide how much time can be allocated to assessing and supporting each individual student (Herppich et al., 2018). In other words, this design of the assessment process supports a high functional correspondence between the tasks in the video-based simulation and real classroom situations, which is accompanied by a physical resemblance that emerges from the use of videos. Altogether, this underpins that the environment can serve as an approximation of practice. Further, the use of scripted videos with a specific focus on the assessment task also supports the simulation as a decomposition of practice (Codreanu et al., 2020; Grossman et al., 2009). This is also supported by the possibility of taking notes during and after observing the simulated students in the videos (see Figure 5). In these notes, participants are asked to coherently describe relevant events in the video, interpret and explain these events, and make conclusions about further learning (Seidel & Stürmer, 2014).

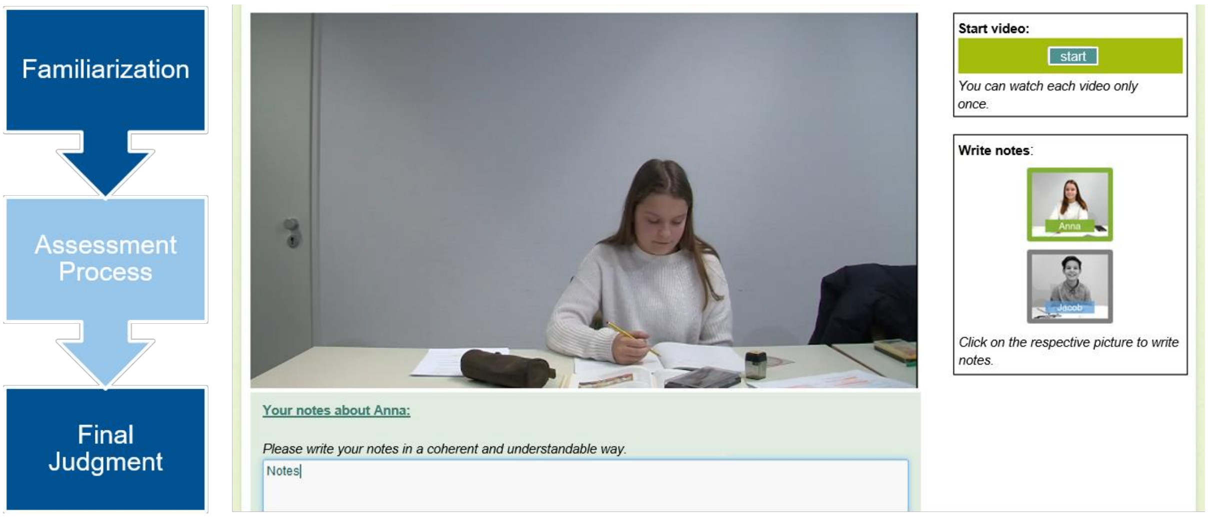

Figure 5. Observation of the video and taking notes during the assessment process.

[7] Note that four simulated students were recorded in total, of which two were used in Study A and in the pre-test of Study B, whereas the other two students were used in the main training session of Study B to reduce training effects.

If the participants think that they have gathered enough information about both simulated students or have reached the maximum number of 10 videos in total, they enter the *final judgment* part of the simulation. At this stage, participants are first asked to formulate a written judgment about each of the two simulated student's mathematical proof skills. Afterwards, they are asked to rate the simulated students' mathematical proof skills on a Likert rating (see Figure 6 and Codreanu et al., 2020). From this rating, the participants' judgment accuracy is obtained (Schrader, 1989).

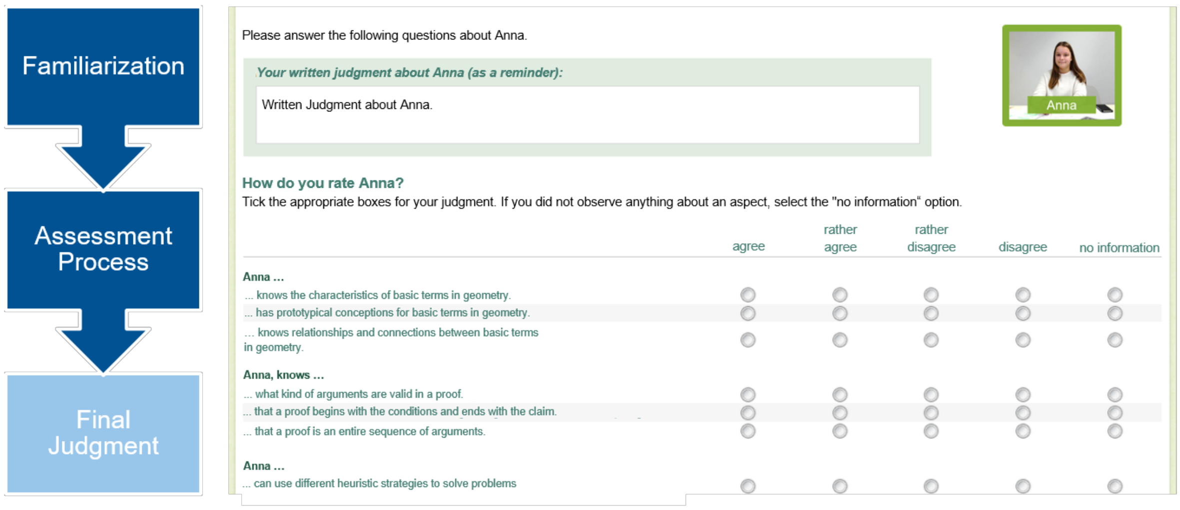

Figure 6. Rating a student's mathematical proof skills during the final judgment.

4.3 Scaffolds

The presented simulation can be used to assess teachers' assessment skills (*Visit-Math Assess*). It can also be used as a learning environment (*Visit-Math Learn*). In the latter case, additional scaffolding can be implemented to increase possible learning gains. Conceptual prompts as conceptual scaffolding were implemented in a previous study (Sommerhoff et al., 2023) and were also implemented in Study B. These prompts were shown in advance of every video and guided the participants' attention to a given aspect of the video and their interpretation of the corresponding indicator for mathematical proof skills. Therefore, they were expected to reduce the complexity of the given assessment task and provide a structured way for noticing and interpreting assessment-relevant aspects in the video (Seidel & Stürmer, 2014). Following these ideas, each prompt was adapted to a specific video during the assessment process according to the following generic example:

> *Pay special attention to* [a relevant aspect in the specific video].
> *What can you conclude about* [a corresponding indicator of the students' proof skills]?

Example: *Pay special attention to how the student explains a parallelogram. What can you conclude about the student's problem-solving strategies?*

In addition to these conceptual prompts, a utility value intervention was newly designed as motivational scaffolding. The utility value intervention was implemented prior to the familiarization part of the simulation. Contentwise, the designed utility value intervention drew on the combination of the two strategies of directly *communicating* and *self-generating* utility value, which has been shown to be effective in prior research (Canning & Harackiewicz, 2015; Gaspard et al., 2021). Thus, the utility value intervention consisted of two parts. In the first part, a five-minute video was shown, in which an experienced researcher explained why assessment skills are particularly relevant for teaching practice. In the video, the experienced researcher gave examples of different assessment situations in teachers' professional lives and explained that accurate assessment can lead to more appropriate individual support and higher instructional quality (Helmke & Schrader, 1987; Nückles et al., 2005). Additionally, statements of future teachers about the relevance of assessment skills were included in this video. In the second part, participants were then asked to explain why it is particularly important and useful for them and their future professional lives to be able to assess students adequately. The thumbnails of both scaffolds are depicted in Figure 7.

(a) Conceptual prompts (b) Utility value intervention

Figure 7. Thumbnails of the implementation of both scaffolds.

4.4 Design of the Studies

4.4.1 Course of Study

Each overarching research question (1 and 2) was addressed in one experimental study. The first study aimed to answer Research Question 1 and took about 90 minutes. Participants were first surveyed regarding their CK and PCK and then asked to self-report their interest, self-efficacy, and self-regulation. Then, they entered the video-based simulation, in which they had to assess the students' mathematical proof skills. In the simulation, dependent variables regarding the assessment process and judgment quality were obtained. After watching the first video, the dependent variables utility value and success expectancy were measured using self-reports. After the simulation, participants were asked to self-report the remaining dependent variables regarding situative learning experiences (presence, authenticity, intrinsic cognitive load, extraneous cognitive load, and germane resources). At the end, demographic data were obtained. The course of this study is depicted in Figure 8.

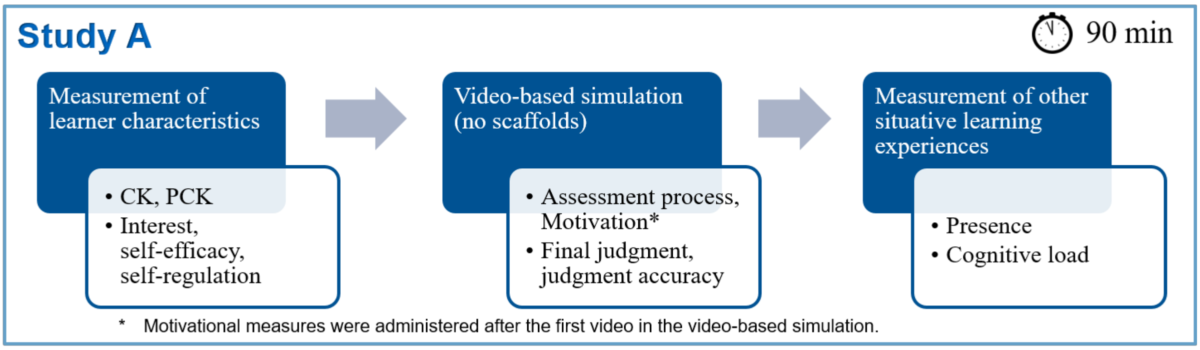

Figure 8. Design of Study A.

Study B focused on the effectiveness of different scaffolding strategies. It consisted of a pre-test and a main training session; the latter was administered at least seven days after the pre-test. Both study parts have a parallel structure. In the pre-test, participants completed a video-based simulation in which the dependent variable judgment accuracy was measured and reported dependent situative learning experiences (utility value, intrinsic cognitive load, and extraneous cognitive load) and their success expectancy (independent variable). In the main training session, the participants again completed the video-based simulation and assessed two other simulated students. This time, the assessment process was scaffolded. Participants were randomly assigned to one of four conditions of a 2 × 2 intervention design: They either received only the utility value intervention (UVI condition), only the conceptual prompts (CP condition), both scaffolds (UVI+CP condition), or no additional scaffolding (control condition). The control group did not receive any additional scaffolds compared to the pre-test. The UVI group received the utility value intervention described above but no conceptual prompts. The CP condition received conceptual prompts during the videos but did not receive an additional utility value intervention. The UVI+CP condition received both the utility value intervention before the simulation started and the conceptual prompts during the videos. For comparable results and as opposed to the pre-test, the number of videos, and thus the number of conceptual prompts for CP and UVI+CP conditions, was fixed to eight videos/ conceptual prompts in the main training session (four videos/conceptual prompts per simulated student). Furthermore, to simulate the writing task of the UVI and UVI+CP conditions, the control group and the CP group were asked to write a short essay about the role of feedback in learning. After the scaffolded assessment process, the dependent variable judgment accuracy was measured. Afterwards, the other dependent variables were also measured (utility value, intrinsic cognitive load, and extraneous cognitive). The course of Study B is shown in Figure 9.

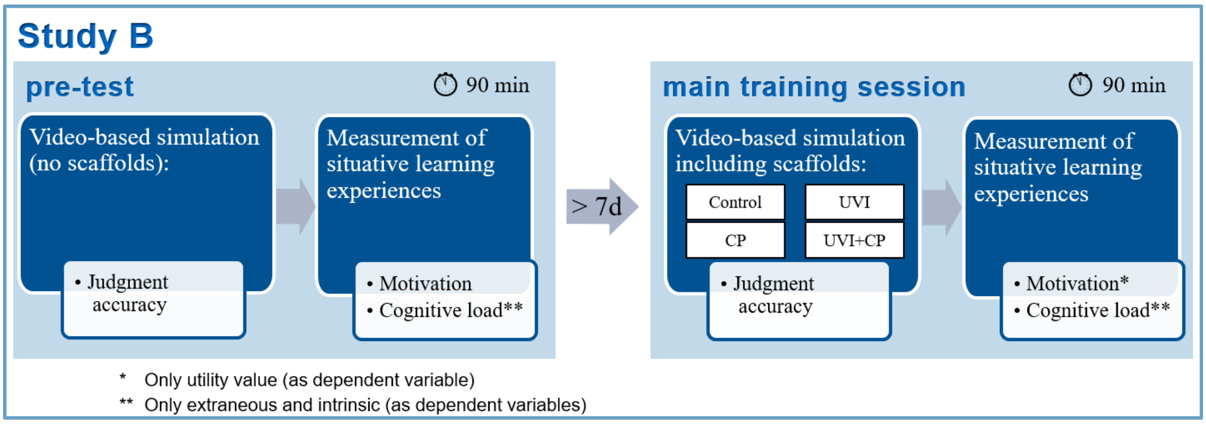

Figure 9. Course of Study B.

4.4.2 Measures and Instruments

In Study A, participants were asked to take a test to measure their CK and PCK. This test consisted of open questions, true-false items, and order tasks. The test was developed by the project team and used in a previous study (Sommerhoff et al., 2023). The participants also self-reported their motivational learner characteristics (interest, self-efficacy, and self-regulation) on scales adapted from prior research (see Table 1). The participants' situative learning experiences were also measured using self-reports with scales adapted from prior research (details can be found in Table 1). Except for the scale on germane resources (Cronbach's α = .65), the internal consistency of these self-report scales ranged from acceptable to excellent (see Article A for details).

Regarding the main outcome measure, assessment skills, measures regarding the assessment process, and judgment quality were used. To describe the assessment process, the total number of videos a participant watched and the total number of words a participant wrote in his or her notes were considered. To describe the written judgment, the total number of words a participant wrote in the final judgments regarding both simulated students was considered. As the main measure of the judgments' quality, judgment accuracy was used (see Figure 6). As mentioned in the description of the simulation, the participants rated the simulated students' mathematical proof skills on a Likert rating at the end of the simulation. More precisely, participants rated eight statements regarding mathematical proof skills for each of the two simulated students on a four-point Likert scale (from 1: "I disagree" to 4: "I agree"). This rating was then compared to an expert solution; only a match with the expert solution was scored with one point, and no match resulted in zero points.

Construct	Scale	Adapted from
Motivational learner characteristics		
Interest	9 items on a 5-point Likert scale	Rotgans and Schmidt (2014)
Self-efficacy	6 items on a 5-point Likert scale	Kunter et al. (2002)
Self-regulation	6 items on a 4-point Likert scale	Kunter et al. (2016)
Situative learning experiences		
Success expectancy	4 items on a 7-point Likert scale	Rheinberg et al. (2001)
Utility value	4 items on a 4-point Likert scale	Wigfield (1994)
Presence	4 items on a 5-point Likert scale	Seidel et al. (2011), Frank (2015), Vorderer et al. (2004)
Authenticity	6 items on a 5-point Likert scale	Seidel et al. (2010), Schubert et al. (2001)
Extraneous cogn. load	3 items on a 7-point Likert scale	Klepsch et al. (2017)
Intrinsic cogn. load	2 items on a 7-point Likert scale	Klepsch et al. (2017)
Germane resources	2 items on a 7-point Likert scale	Klepsch et al. (2017)

Table 1. Self-report scales in Study A.

In Study B, success expectancy was measured using the same scale as in Study A. To reflect the focus on the utility value intervention, the self-report scale on utility value was extended by adding four items on a 4-point Likert scale adapted from Canning and Harackiewicz (2015). The other dependent self-reported variables (intrinsic and extraneous cognitive load) were measured as in Study A. Except for the scale on extraneous cognitive load in the pre-test (Cronbach's $\alpha = .68$), the internal consistency of these self-report scales ranged from acceptable to good (see Article B for details). The remaining dependent variable, judgment accuracy, was operationalized and calculated as in Study A.

4.5 Data Collection

Both studies were conducted in 2020 and 2021 using an online environment in which the simulation was embedded. Due to the COVID-19 pandemic, both empirical studies had to be conducted online. This allowed participants more flexible access to the simulation, as the simulation was always available, and participants could access it on their own devices at home. To recruit participants, the simulation was promoted in university courses. In these courses, future teachers were invited to voluntarily participate in the studies as an opportunity to gain further understanding of the courses' learning goals. In one lecture, it was possible to embed the main training session of Study B. In this case, participation in the study (i.e., by completing the pre-test in advance and allowing the use of anonymized data for research purposes) was optional but remunerated. Note that for all participants, participation was remunerated and did not influence the participants' grading. Both studies were approved by the Technische Universität München (TUM) data protection office and the ethical committee of the German Psychological Society (DGPs).

For Study A, data were collected from 150 future teachers from different German universities. These future teachers were mainly female, which is typical of German teacher education (101 female [67.3%], 47 male [31.3%], and 2 NA [1.3%]). As the standard study time for future teachers at German universities is nine to ten semesters, the sample was in the middle of their studies ($M_{\text{semester}} = 5.0; SD_{\text{semester}} = 2.5$). For Study B, data were collected from 108 future teachers from different German universities. Again, the sample was predominantly female [76 female (70.4%), 31 male (28.7%), and 1 NA (0.9%)]. In this study, future teachers were slightly more advanced in their studies than in Study A ($\mathrm{M}_{semester} = 6.4; SD_{semester} = 3.2$).

4.6 Data Analysis

To analyze the collected data, the software R (R Core Team, 2020) and Mplus (Muthén & Muthén, 1998-2017) were used. More precisely, pre-processing of the data (e.g., calculating mean scores of the scales) and descriptive statistics was done using R.

For Article A, data were then transferred to Mplus using the MplusAutomation package (Hallquist & Wiley, 2018). In Mplus, LPA was conducted using the measured learner characteristics as indicators (CK, PCK, interest, self-efficacy, and self-regulation). The Akaike information criterion (AIC), the Bayesian information criterion (BIC), the adjusted BIC (aBIC), the Lo-Mendell-Rubin test (LMR), and the Bootstrapped Likelihood Ratio Test (BLRT) were used to determine whether and how many distinct latent profiles could be identified (Spurk et al., 2020). To ensure that these profiles were sufficiently distinct, entropy and posterior classification probabilities were also considered. Using these indicators and interpretability as criteria, one- to five-profile solutions were compared to analyze the presence of distinct learner characteristic profiles and the number of these profiles, if present. This represents the first step of the commonly used three-step approach for conducting LPA (Vermunt, 2010). Next, as learner characteristic profiles could be identified, participants were assigned to these latent profiles. It was drawn on an approach that assigns participants to the profile for which membership is most likely based on the posterior classification probability ("modal assignment") (see Vermunt, 2010 for details and other assignment strategies). In the last step, the means of these profiles were calculated regarding several outcome variables (here: situative learning experiences and assessment skills). In this step, calculating unweighted profile means for outcome variables would have been biased, as participants could not be *uniquely* assigned to one profile in most cases. To more generally account for the fact that a participant's predicted latent profile is not necessarily the true latent profile, the BCH method was developed (Bolck et al., 2004; Dziak et al., 2016; Vermunt, 2010). Thus, for the current data analysis, the BCH method was used to calculate profile means regarding situative learning experiences and assessment skills. Wald's χ^2 test was used to examine whether these means differed (Spurk et al., 2020).

For Article B, analyses of variance (ANOVAs) were calculated to check whether the different conditions differed in the dependent variables (utility value, intrinsic cognitive load, extraneous cognitive load, and judgment accuracy) in the pre-test. To compare the development from pre- to post-test[8] regarding the dependent variables, two-way repeated measures ANOVAs were calculated (dependent variable: judgment accuracy; independent variables: condition, time). Moreover, the pre- and post-test values were compared for each single condition using *t*-tests. To test for possible differential effectiveness, a linear mixed model was calculated using the lmer() function from the lme4 package (Bates et al., 2015). Subsequently, by setting the learners' success expectancy to one standard deviation above/below mean in this model, the marginal means for learners with a comparably high/low success expectancy were estimated and compared using the emmeans package (Lenth, 2022).

Note that both articles draw on an alpha level of α = .05 to test statistical significance. Further details regarding the statistical analysis can be found in both articles.

[8] As in article B, *post-test* values refer to all measures that were obtained after the scaffolded assessment process, but still within the main training session.

5 Summary of Publications

5.1 Article A: *Video-Based Simulations in Teacher Education: The Role of Learner Characteristics as Capacities for Positive Learning Experiences and High Performance*

This first article addresses the superordinate Research Question (1) by answering the related Research Questions 1.1, 1.2, and 1.3 (see Section 3.1). It investigates whether patterns of learner characteristics that future teachers bring to the video-based simulation can be identified. To analyze whether the indicators theoretically derived by Heitzmann et al. (2019) can serve as a basis for additional support through scaffolding, the connection between identified learner characteristic profiles and situative learning experiences, as well as with the assessment skills, were examined (Brown et al., 1982; Friedrich & Mandl, 1992).

With regard to Research Question 1.1, LPA was used to describe patterns of co-occurring scores across the indicators CK, PCK, interest, self-efficacy, and self-regulation. Based on fit indices and interpretability, three learner characteristic profiles were identified: The *knowledgeable* profile including about 41% of the total sample size showed above-average CK and PCK but average scores regarding motivational learner characteristics (interest, self-efficacy, and self-regulation). The *motivated* profile comprised about 25% of the total sample and had above-average motivational characteristics but average CK and slightly below-average PCK. The scores of the last profile were below average and below both other profiles across all indicators. Thus, this profile was called *potentially struggling*, and about 35% of the total sample was allocated to this profile.

To explore how these profiles were connected to situative learning experiences and assessment skills, we used the BCH method to calculate the profile's means, and compared them using a Wald's χ^2 test. With regard to Research Question 1.2, significant results revealed that the *motivated* profile was connected to motivational situative learning experiences through a higher success expectancy component of motivation. The *knowledgeable* profile, in turn, was connected to cognitive situative learning experiences through a higher use of germane resources. However, the profiles did not differ regarding authenticity and presence.

Regarding Research Question 1.3, the results indicated that future teachers with a *knowledgeable* profile watched significantly more videos and also reached significantly higher judgment accuracy than future teachers with a *potentially struggling* profile. Conversely, future teachers with a *motivated* profile did not show significant (dis)advantages compared to future teachers with a *potentially struggling* profile or a *knowledgeable* profile. Interestingly, future teachers with a *motivated* profile tended to show an increase in the number of written words in the final judgment compared to the number of written words in the notes during the assessment process, whereas both other profiles wrote approximately the same number of words as during the assessment process.

In summary, the present study used a person-centered approach to analyze the role of patterns of future teachers' learner characteristics in learning assessment skills in a video-based simulation (Woo et al., 2018). The results underpin the important role of knowledge in assessment skills by pointing toward the advantages of future teachers with a *knowledgeable* profile in their judgment compared to both other profiles (Herppich et al., 2018; Loibl et al., 2020). In particular, the advantages of future teachers in the *knowledgeable* profile over those in the *motivated* profile point to a major role of knowledge compared to motivational characteristics with regard to assessment skills, which has also been found in other fields (Hattie, 2009). Yet, the results also potentially point toward the benefits of future teachers with the *motivated* profile, which stands out for its above-average motivational characteristics, with regard to assessment skills (compared to future teachers with the *potentially struggling* profile). By pointing toward the relevance of motivational characteristics for motivational situative learning experiences and the relevance of cognitive characteristics for cognitive situative learning experiences, the results also contribute to further validation of the framework of Heitzmann et al. (2019).

On a more general level, these results indicate that future teachers with different learner characteristic profiles may differ in their situative learning experiences as well as in their assessment skills (Heitzmann et al., 2019). Thus, it seems reasonable to base scaffolding strategies on these profiles (Plass & Pawar, 2020; Tetzlaff et al., 2021). In that regard, this article indicates that there are *knowledgeable* learners who may benefit most from motivational support, such as a utility value intervention, as they already possess above-average knowledge. Similarly, there are also *motivated* learners who may benefit from conceptual prompts as cognitive support. Lastly, learners from the third profile, who showed below-average knowledge and motivational characteristics, may benefit mostly from a combination of a utility value intervention and conceptual prompts.

5.2 Article B: *Fostering Pre-Service Teachers' Assessment Skills in a Video Simulation: Differential Effects of a Utility Value Intervention and Conceptual Knowledge Prompts*

Building on the finding of Article A that different learners may benefit differently from cognitive and/or motivational support for answering Research Question (2), Article B aimed at investigating whether conceptual prompts and a utility value intervention as promising representatives for cognitive and motivational support can be effectively used to foster teachers' assessment skills (Hulleman & Harackiewicz, 2021; Sommerhoff et al., 2023). As *potentially struggling* future teachers may require both kinds of support, it also investigated whether a combination of a utility value intervention and conceptual prompts is beneficial for teachers' assessment skills, which could be expected from research in other domains (Cromley et al., 2020). This led to a 2 × 2 intervention design (conditions: Control, UVI, CP, and UVI+CP; see also Section 4.4.1), in which the effectiveness of the scaffolds was evaluated with regard to future teachers' judgment accuracy as the main indicator of assessment skills (Urhahne & Wijnia, 2021). Situative learning experiences were also measured (extraneous and intrinsic cognitive load, and utility value; see also Figure 3) to investigate possible effect mechanisms.

An analysis of the pre-test data revealed that the four conditions did not differ significantly in the pre-test regarding the dependent variables. With regard to judgment accuracy, all groups improved descriptively from pre- to post-test, but only the CP condition improved significantly. In particular, the improvement in the UVI condition did not reach significance, even though this condition and the CP condition had descriptively similar judgment accuracy in the post-test. Surprisingly, future teachers receiving both types of scaffolds did not significantly improve in their judgment accuracy, and they even had the lowest descriptive improvement compared to the other conditions.

Regarding situative learning experiences, it is important to note that future teachers' perception of the utility value was already almost at the top of the scale in the pre-test, which could have led to a ceiling effect. Regarding other situative learning experiences, only the control group perceived significantly less extraneous load during the intervention than in the pre-test. No other changes within future teachers' situative learning experiences reached significance. The decreases in the perception of intrinsic cognitive load of groups receiving conceptual prompts (CP and UVI+CP) tended to yield small to medium effect sizes.

In summary, these results substantiated prior findings in the context of *Visit-Math* (Sommerhoff et al., 2023) that conceptual prompts are a promising means of facilitating future teachers' assessment skills. Interestingly, when combined with a utility value intervention, conceptual prompts did not yield further significant benefits as opposed to prior attempts to combine cognitive and motivational support (Cromley et al., 2020). When the utility value intervention was presented without conceptual prompts, there was only a descriptive tendency for the utility value intervention to be effective in facilitating future teachers' assessment skills. This was not expected, as prior research pointed toward the effectiveness of utility value interventions, yet with rather small effect sizes (Hulleman & Harackiewicz, 2021). In that regard, previous research has indicated that utility value intervention can be especially effective for learners with comparably high success expectancy (Durik et al., 2015). This was also found in the present study, as future teachers with comparably high success expectancy tended to benefit most from the utility value intervention, classified with a medium to large effect size for these learners. All other conditions (control, CP, and UVI+CP) had only small or small to medium effect sizes. Conversely, for learners with a success expectancy of one standard deviation below the mean, conceptual prompts were most helpful, showing a large effect size, possibly indicating a motivational effect of conceptual prompts (Belland et al., 2013). Thus, both scaffolds seem to be differentially effective for learners with comparably high or comparably low success expectancy.

6 Discussion

6.1 Discussion of Central Findings

6.1.1 Role of Learner Characteristics in Performance and Learning

One of the goals of Article A was to identify typical patterns of cognitive and motivational learner characteristics among future teachers. By using LPA, it was possible to identify three such patterns: a *knowledgeable* profile, a *motivated* profile, and a *potentially struggling* profile. Each profile had similar scores across motivational learner characteristics (e.g., no profile had clearly above-average interest but clearly below-average self-efficacy) and similar scores across aspects of professional knowledge (e.g., no profile had clearly above-average CK but clearly below-average PCK). Thus, for better interpretability, it makes sense to summarize profiles by combining mean values of the profiles to one mean value for cognitive characteristics (CK, PCK), and one mean value for motivational characteristics (interest, self-efficacy, self-regulation); see Figure 10.

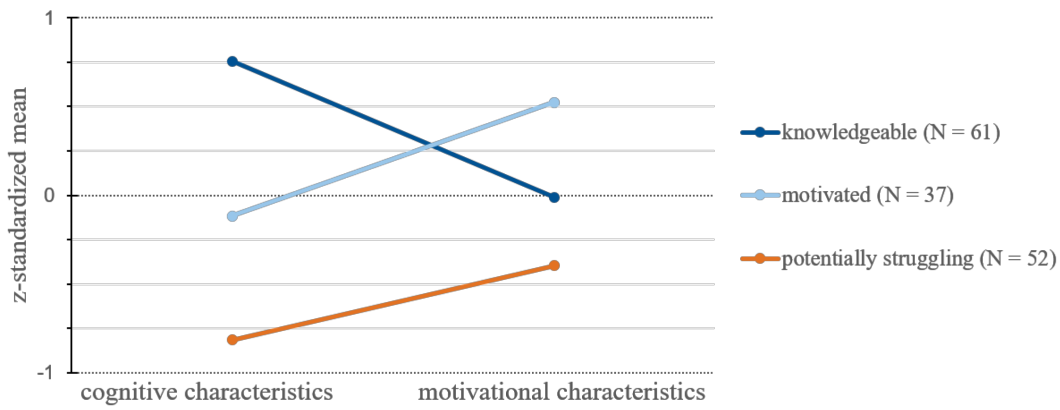

Figure 10. Summary of the three identified latent profiles.

The three profiles met the expectations: profiles with above-average knowledge but below-average motivational characteristics and vice versa, as well as a profile with both below-average knowledge and below-average motivational characteristics. However, a fourth profile with above-average knowledge and above-average motivational characteristics was also expected but could not be detected. Note that this does not mean that there are no future teachers with above-average professional knowledge and above-average motivational characteristics. However, this combination could not be identified as a typical pattern of future teachers' learner characteristics and—at least in the sample of Article A—the combination of above-average professional knowledge and above-average motivational characteristics was scarce. This missing fourth profile seems particularly important, as both professional knowledge and motivational characteristics are regarded as crucial aspects of teachers' professional competence (Baumert & Kunter, 2006; Kunter, Klusmann, et al., 2013). Thus, further research is needed to investigate whether this promising fourth profile can be detected in other samples, for example, in samples of more experienced teachers. With regard to the three identified profiles, future teachers with a *knowledgeable* profile can be considered learners who met the university's primary learning goal of acquiring knowledge *most* compared to future teachers with either of the other two profiles (Kunter, Kleickmann, et al., 2013). Fortunately, as an empirical validation of universities' focus on knowledge in teacher education, future teachers with this profile had descriptively the highest judgment accuracy and reported the most use of germane cognitive resources—in both cases, with significant advantages over the *potentially struggling* future teachers.

Due to its advantages in judgment accuracy and cognitive situative learning experiences, the *knowledgeable* profile may serve as a role model profile (Schrader, 1989; Sweller et al., 2019). Since future teachers with a *knowledgeable* profile are more accurate in their judgments than future teachers with either of the other two profiles, it can be assumed that they also draw on a better assessment process (Brandl et al., 2021; Herppich et al., 2018). When considering their assessment process, advantages could be reflected in (i) the notes future teachers' wrote during the assessment process and in (ii) the number of videos future teachers watched.

With regard to notes (i), future teachers with a *knowledgeable* profile wrote most words in their notes overall. Since the total sum of words also depends on the total number of watched videos, it seems reasonable to take a particular look at the average number of words written per video (see Table 2). This allows to investigate the quantity of the notes more independently from the quantity of videos watched.

Construct	**Knowledgeable M (SE)**	**Motivated M (SE)**	**Potentially struggling M (SE)**
Number of words in the notes per video	19.83 (1.70)	19.77 (2.68)	18.02 (1.88)

Table 2. Mean and standard error of the number of words the three profiles wrote in their notes per watched video.

These calculations highlight that the three profiles did not vary considerably in the number of words written in their notes (χ^2 = 0.55, *p* = .760). Thus, the reason for a higher quality in the assessment process of future teachers with a *knowledgeable* profile may not be the quantity of their notes but the quality. That is, the notes may better mirror the relevant video events that future teachers notice and show better descriptions, explanations, and predictions (Brunswik, 1955; Herppich et al., 2018; Seidel & Stürmer, 2014). A first approach that provides insights into the quality of the written notes is to consider the extent to which the three indicators of students' mathematical proof skills are covered (Codreanu et al., 2021; Sommerhoff, 2017). The first results indicate that the future teachers with a *knowledgeable* profile covered basic content knowledge and methodological knowledge to a greater extent than those with a *potentially struggling* profile, whereas problem-solving strategies were considered to a similar extent among all profiles (Nickl, Sommerhoff, Codreanu, et al., in press). This substantiates that the advantages of future teachers with a *knowledgeable* profile may be mirrored in the *quality* of their notes.

These advantages may also be mirrored in the number of watched videos (ii), since *knowledgeable* future teachers watched significantly more videos compared to *potentially struggling* ones, which may have allowed them to notice more relevant video events (Brunswik, 1955; Codreanu et al., 2022). However, it is worth noting that watching more videos could decrease the efficiency of the resulting judgment (Heitzmann et al., 2019). This is of particular importance within classroom settings when teachers have to attend to the whole class while supporting individual students. Spending too much time with the assessment of a single student in such situations could decrease the accuracy of judgments regarding other students; see Braun et al. (2017) for a similar discussion in medical education. As Braun et al. (2017) pointed out, efficiency is about finding a balance between adequate levels of judgment accuracy and coming to judgments quickly. However, in the context of acquiring assessment skills, it seems reasonable to first reach a sufficient level of judgment accuracy before increasing the judgments' efficiency. Even though there are no clear thresholds for a sufficient level of judgment accuracy (Blömeke & Kaiser, 2017), future teachers' judgments matching experts' judgments less than 50% on average, which was the case in Article A and Article B, and also holds for the *knowledgeable* profile, suggests that improving judgment accuracy still needs to be prioritized. Thus, despite a possible decrease in efficiency, watching more videos can be regarded as a positive characteristic of future teachers' assessment process – at least when their judgment accuracy is on a lower level, as watching more videos possibly allows them to notice more relevant events in the videos and may thus increase judgment accuracy.

Thus far, the discussion has mainly focused on the assessment performance of future teachers with a *knowledgeable* profile in comparison to those with a *potentially struggling* profile. This has underpinned the important role of knowledge in future teachers' assessment skills, which is also in line with prior research (Kramer, Förtsch, Boone, et al., 2021; Todorova et al., 2017). Further, the joint approach to analyzing cognitive and motivational characteristics seemed promising to further investigate the role of motivational characteristics in teachers' assessment skills. However, the results are inconclusive in this regard. The assessment process and judgment accuracy of future teachers with a *motivated* profile did not differ significantly from those with a *knowledgeable* profile. Yet, there are also no significant differences to future teachers with a *potentially struggling* profile, making the *motivated* profile an "in-between" profile and making possible interpretations regarding the *motivated* profile tentative. That is, the *motivated* profile may not be on par with the *knowledgeable* profile regarding assessment skills, but it also seems better than the *potentially struggling* profile. Thus, the motivational advantages of future teachers with a *motivated* profile may not be able to completely compensate the advantages in knowledge of future teachers with a *knowledgeable* profile. This would also be in line with findings from other learning contexts, in which knowledge often had a greater impact on performance than motivational characteristics (e.g., Hattie, 2009).

One possible driver that might have affected the assessment performance of future teachers with the *motivated* profile negatively compared to those in the *knowledgeable profile* may lie in their above-average self-efficacy. For example, Klug et al. (2016) found self-efficacy to negatively predict assessment performance and discussed this finding against the backdrop of possible overestimation. Similarly, manifested in comparably high self-efficacy and success expectancy, future teachers with a *motivated* profile were comparably optimistic about assessing the simulated students accurately, which they, in fact, did not (by matching only about 33% with the experts' judgment), which points to a possible overestimation of future teachers with a *motivated* profile. Even though the findings by Klug et al. (2016) point to a critical role of overestimation for assessment skills, and other studies, such as Robins and Beer (2001), also found a connection between overestimation and disengagement, it is important to note that, in general, overestimation has been reported as rather beneficial for learning (Dupeyrat et al., 2011; Praetorius et al., 2016). Thus, it seems desirable to investigate how overestimating learners, such as those with the *motivated* profile, may be supported in acquiring assessment skills. As one possible approach to address this research aim, the role of success expectancy in the effectiveness of scaffolds was investigated in Article B.

6.1.2 Scaffolding

Article A revealed different patterns of learner characteristics and their relation to situative learning experiences and assessment skills. As a result, possible approaches for adequately fostering assessment skills, such as motivational support and cognitive support, became evident. The main goal of Article B was to investigate the effectiveness of a utility value intervention as an example of motivational support and conceptual prompts as an example of cognitive support using a 2 × 2 intervention design. The potential effectiveness of combining these two strategies was also investigated. The results of the four conditions (control, conceptual prompts only, utility value intervention only, conceptual prompts, and utility value intervention) will be discussed in the following section.

The results indicate that, on average, future teachers in the control group improved descriptively from pre- to post-test in their judgment accuracy. One possible reason for this may be the ongoing familiarization with the simulation, which resulted in a significant decrease in extraneous cognitive load from the pre-test to the post-test for the control group (Sweller et al., 1998). By receiving the utility value intervention or cognitive prompts, all other conditions had to deal with additional instructional elements related to the learning context, making it plausible that these conditions did not change significantly in their extraneous cognitive load, despite their familiarization with the other instructional elements of the simulation (Mayer, 2014). However, regarding judgment accuracy, the descriptive improvements of the control group did not reach significance.

In that regard, the condition receiving only conceptual prompts improved significantly from the pre-test to the post-test, which is in line with prior research drawing on the same conceptual prompts (Sommerhoff et al., 2023). Thus, this study further substantiates that conceptual prompts can be effectively used to foster assessment skills in simulations and may also have a strong impact, as, similarly to Sommerhoff et al. (2023), who observed strong effects of conceptual prompts, Article B reported a medium to strong effect. The magnitude of this effect size is slightly higher than what was reported in a meta-analysis by Chernikova, Heitzmann, Fink, et al. (2020), who analyzed the effectiveness of prompts in simulations within teacher education and medical education. With regard to the effect mechanism of the conceptual prompts from Article B, learners receiving conceptual prompts reported a small- to medium-sized tentative decrease in intrinsic cognitive load. This suggests that the designed conceptual prompts seemed to have influenced the intrinsic cognitive load, which in turn might have led to an increase in judgment accuracy—as expected (Klepsch & Seufert, 2021). Prior research has also pointed toward the possible motivational effects of conceptual prompts. For example, conceptual prompts have been shown to increase learners' success expectancy (Bixler, 2007), which could also be the case in Study B, as the conceptual prompts in Study B may have been perceived as additional help and thus led to more optimism for assessing the simulated student accurately. In other words, conceptual prompts may have positively affected future teachers' success expectancy component of motivation, and may have thus increased their performance (Wigfield & Eccles, 2000). This is empirically substantiated by the finding of Article B that learners with a comparably low expectancy benefitted strongly from conceptual prompts, whereas learners with one standard deviation above the mean (representing a success expectancy of 4.71 + 1.12 = 5.83 on the unstandardized 7-point Likert scale) benefitted less from this expectancy boost, possibly due to a ceiling effect. This is particularly interesting, as prior analyses revealed that future teachers' knowledge does not influence the effectiveness of these conceptual prompts, and conceptual prompts may thus be not particularly beneficial for a certain learner subgroup but generally effective (Sommerhoff et al., 2023). In that regard, the findings of Article B highlight that there are variables beyond knowledge, such as motivational characteristics, that may influence the effectiveness of conceptual prompts (Belland et al., 2013). Figure 11 depicts the estimated effect sizes for conceptual prompts (turquoise line) and for the other conditions in Study B plotted over the learners' success expectancy (calculated in the same way as in Article B). In particular, this figure suggests that future teachers with a success expectancy of up to M+0.60 SD (on the unstandardized scale from 1 to 7, this corresponds to a value of 5.38; $N = 72$ future teachers of the sample in Study B had a lower success expectancy) may have benefitted most from conceptual prompts. Learners with higher success expectancy, in turn, may have benefitted most from the utility value intervention.

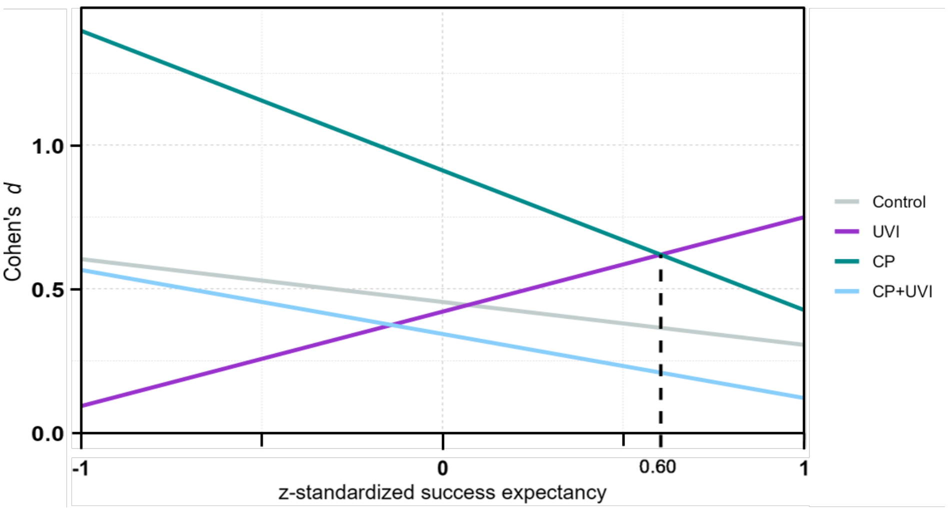

Figure 11. Effect size of the four conditions plotted over future teachers' success expectancy.

With regard to the measured utility value, it is surprising that future teachers considered assessment skills an important component of their later professional lives, since they typically lack a perception of relevance regarding learning contents within educational sciences (Alles et al., 2019; Voss, 2022). This could be caused by a pre-selection of motivated future teachers, as participation was voluntary yet mitigated through remuneration (see Section 6.4). However, future teachers' high perception of utility could also point toward a motivational effect of the video-based simulation. As the utility value was measured after future teachers completed the video-based simulation in the pre-test of Study B, the simulation itself could have increased future teachers' utility value regarding assessment skills by explicitly showing the relevance of assessment skills in an application case (Belland et al., 2013). In the end, future teachers' increased perception of utility value prior to the actual utility value intervention may have led to a ceiling effect of the intervention, resulting in a descriptive yet nonsignificant increase in judgment accuracy for future teachers receiving the utility value intervention. However, there are also other aspects of motivation, such as success expectancy, involvement, or interest, that may have been affected by the utility value intervention and could have led to better performance and learning outcomes. The detection of effects on performance and learning outcomes, however, may be challenging, as Hulleman and Harackiewicz (2021) found the effect sizes of utility value interventions regarding learning outcomes to be rather small ($d = .24$). Thus, as the sample size of Article B only allowed the detection of small to medium effect sizes, the utility value intervention might have been effective, even though the descriptive improvement did not reach significance. This does not seem unrealistic, as the future teachers who received the utility value intervention had at least descriptively higher judgment accuracy in the post-test than the control condition ($t(54) = .858$, $p = 0.395$, $d = .23$). Against the backdrop of these small effect sizes, further research may also consider other types of motivational scaffolding, as, for example, competence-related beliefs have been shown to have more impact on performance and learning gains compared to the utility value (Rosenzweig et al., 2022). However, prior research has repeatedly stressed that utility value interventions are not one-size-fits-all interventions but yield higher effect sizes for certain learner subgroups (Hulleman & Harackiewicz, 2021). As such, prior research has focused on the moderating role of success expectancy, which was also used for analyzing the differential effects of utility value interventions in Article B (Canning & Harackiewicz, 2015; Durik et al., 2015). It was found that the designed utility value intervention was most beneficial for learners with a comparably high success expectancy (see also Figure 11). In other words, increasing the utility value as part of the value component of motivation was beneficial for learners with comparably high success expectancy. Thus, in the present discussion on the interaction of value and success expectancy, the results of Article B point to the existence of a positive interaction of both components (Nagengast et al., 2011; Wigfield & Eccles, 2020). This is also in line with the empirical results of Trautwein et al. (2012), who additionally found comparably low success expectancy when combined with comparably high utility value (in other words, knowing about the relevance of a task but not feeling able to cope with it) to be more detrimental than when combined with comparably low utility value. This may be reflected in the finding of Article B that future teachers with comparably low success expectancy improved the least compared to all other conditions when receiving a utility value intervention. Thus, the empirical findings underpin that utility value interventions can be particularly effective for learners with comparably high success expectancy.

Surprisingly, the combination of both scaffolds did not significantly improve the participants' judgment accuracy. Descriptively, the improvement of this condition was even the lowest among all four conditions, which means that combining both scaffolds may have reduced the individual scaffolds' effectiveness and was even less effective than only getting acquainted with the simulation, as in the control group. This is surprising, as the effects of both scaffolds should be independent from each other, since the utility value intervention primarily aims at increasing the perceived utility value and conceptual prompts rather focus on reducing cognitive load and guiding future teachers' attention and conclusions (Hulleman & Harackiewicz, 2021; Klepsch & Seufert, 2021; Sommerhoff et al., 2023) (see also Figure 3). The benefits of combining motivational and cognitive support have also been substantiated by prior research theoretically (e.g., in the context of self-regulated learning, see Pintrich, (2000) and empirically (Cromley et al., 2020). Cromley et al. (2020) combined various types of cognitive and motivational support for undergraduate STEM students, which led to larger effects than cognitive-only or motivational-only support. However, they also found that different combinations of motivational and cognitive support varied in their effectiveness. Thus, the less effective combination of motivational and cognitive support in Article B may indicate that utility value interventions and conceptual prompts could be less fruitfully combined, which needs to be substantiated in further research. As an alternative type of cognitive support that may be combined more fruitfully with utility value interventions, self-reflection prompts have shown effective in other assessment contexts (Mamede & Schmidt, 2017), as well as in combination with other types of support (Huk & Ludwigs, 2009). Thus, this may be particularly effective for learners who benefit from both cognitive and motivational support, such as future teachers with a *potentially struggling* profile in Article A.

6.2 Overarching discussion: Relation to Adaptivity

Article A highlighted the presence of distinct learner characteristic profiles, which served as a basis for developing conceptual prompts and a utility value intervention as scaffolding measures, which were shown to be potentially helpful when empirically evaluated in Article B. Besides addressing learner characteristic profiles in the *design process* of possible scaffolding measures, prior research has repeatedly pointed toward the benefits of individualized support in the *learning process*, often referred to as adaptive support (Kulik et al., 1990). In the context of technology-enhanced learning, adaptivity is typically referred to as the ability of a learning system to specifically support individual learners (Aleven et al., 2017; Plass & Pawar, 2020). To this end, the theoretical considerations and empirical studies in this dissertation can also be interpreted as an approach for preparing adaptivity within the Visit-Math simulation. This approach is based on two guiding questions: *What variable should the learning environment adapt to? How should the learning environment adapt to these variables* (Aleven et al., 2017; Plass & Pawar, 2020). These two guiding questions will be described in more detail in the following section; see Figure 12 for an overview.

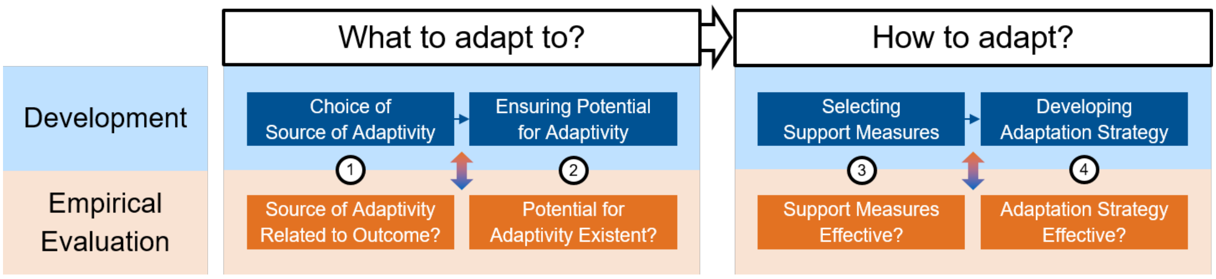

Figure 12. Step-wise approach to establishing adaptivity in a learning environment.

6.2.1 What variable should the learning environment adapt to?

As the first step of this dissertation's approach toward adaptivity, the *source of adaptivity*[9] was identified, which refers to finding variables that can serve as a valid basis for deciding which instruction to present (Aleven et al., 2017; Friedrich & Mandl, 1992; Plass & Pawar, 2020). Generally, the source of adaptivity can be chosen from a great variety of behavioral, cognitive,

[9] The source of adaptivity is sometimes also called *learner model*, for example, in the context of intelligent tutoring systems; see Shute and Zapata-River (2012). See also Brusilovsky and Millán (2007) for an overview of different approaches for modeling the source of adaptivity.

motivational, emotional, and socio-cultural variables (Plass & Pawar, 2020; Vandewaetere et al., 2011), but it is important that the chosen variables are (theoretically and empirically) related to the learning outcome and can be measured adequately (Plass & Pawar, 2020). For this dissertation, Heitzmann et al.'s (2019) theoretical considerations highlighted the relevance of cognitive (professional knowledge) and motivational characteristics (interest, self-efficacy, and self-regulation) for teachers' assessment skills. As opposed to common, solely theoretical approaches to substantiate the selection of the source of adaptivity (Vandewaetere et al., 2011), this dissertation follows the recommendations of Plass and Pawar (2020) by dedicating an empirical study to the validation Heitzmann et al.'s (2019) theoretical considerations. To this end, using LPA reflects that learner characteristics typically appear in complex patterns (Kosel, Wolter, & Seidel, 2021) and methodologically allows the representation of these complex patterns by a single variable (i.e., the latent profile). Whether future teachers are assigned to the *knowledgeable*, *motivated*, or *potentially struggling* profile can also be used as a source of adaptivity, which is backed up by Article A, substantiating the profiles' relation to assessment skills as a learning outcome. Note that the latent profile as a source of adaptivity still draws on the measurement of *multiple* learner characteristics and thus exceeds the common approaches taken in research on adaptivity, which mainly focus on one variable as a source of adaptivity (Nakic et al., 2015). This also emphasizes that LPA can serve as a rather simple and comprehensible approach for including multiple learner characteristics compared to other fairly complex approaches (Kabassi & Virvou, 2004).

Once selected, it must also be ensured that the chosen source of adaptivity bears the potential for adaptivity. That is, the source of adaptivity suggests possible directions of support, and not all learners receive the same support as they would, for example, when varying in a too small extent along the source of adaptivity (see Shute & Zapata-River, 2012). For example, choosing a variable as a source of adaptivity that is already sufficiently present among learners will not suggest possible directions of support and will thus fail to yield a valid adaptation strategy. Again, substantiating theoretical considerations with empirical findings can further ensure the potential for adaptivity (Plass & Pawar, 2020). In this regard, this dissertation underpinned the merits of person-centered approaches, such as LPA, to identify certain learner subgroups for which promising support measures can then be individually developed (Tetzlaff et al., 2021). To this end, Article A highlighted that all profiles could still improve, and future teachers may be supported through cognitive support (*motivated* profile), motivational support (*knowledgeable* profile), or both (*potentially struggling* profile).

6.2.2 How should the learning environment adapt to these variables?

Once the source of adaptivity and its potential for adaptivity have been clarified, the focus shifts to the second major question of *how the learning environment should adapt to these variables*. This question can be disentangled in the selection and development of possible support measures in the learning environment (*how to support*) and in the implementation of the developed support measures in an adaptation strategy (*when to support how*). With regard to the selection of support measures, Plass and Pawar (2020) indicated seven different categories of support while solving a task, such as scaffolding or providing feedback, from which designers of adaptive learning environments can choose. The specific choice of support needs to build upon the chosen source of adaptivity. In the context of the present dissertation, motivational and conceptual scaffolds, as well as their combination, seemed promising to specifically support the identified profiles and were then designed based on theoretical considerations as well as prior empirical results (Hulleman & Harackiewicz, 2021; Sommerhoff et al., 2023). At this point, the present dissertation highlights the benefits of empirically investigating the developed scaffolds' effectiveness in the targeted learning environment, in addition to theoretical considerations (Aleven et al., 2017): Contrary to the expectations, the combination of the utility value intervention and the conceptual prompts tended to be less effective than each scaffold individually. Following these expectations could have led to a less effective use of the scaffolds, particularly for learners who could have benefitted the most from the combination of conceptual and motivational scaffolding (e.g., future teachers with a *potentially struggling* profile), but who may benefit more from receiving only conceptual prompts based on the empirical findings of Study B.

Such empirical results also affect the adaptation strategy, which refers to the question of *when* (in terms of individual scores on the chosen source of adaptivity) the learning environment should support *how* (in terms of the available support measures). Depending on the chosen source of adaptation and the designed support measures, the development of an adaptation strategy can be rather simple. For example, based on the learner characteristic profiles from Article A (as a source of adaptivity) and the utility value intervention and conceptual prompts (as support measures), one possible adaptation strategy could be that, depending on the assigned profile, future teachers receive either conceptual prompts (*motivated* or *potentially struggling* profile) or the utility value intervention (*knowledgeable* profile). However, if the source of adaptivity is not a nominal variable but an ordinal or continuous variable, such as future teachers' success expectancy, thresholds need to be derived for deciding which possible support measure is provided for learners with a certain score on the chosen source of adaptivity (see Shute, 1995). One approach to how such cut-off values could be derived was applied in Section 6.1.2, in which the effectiveness of the utility value intervention and conceptual prompts was compared depending on future teachers' success expectancy (see Figure 11). This points to another possible adaptation strategy, according to which future teachers with a success expectancy higher than 5.38 are provided with a utility value intervention, whereas those with a lower success expectancy than 5.38 receive conceptual prompts. For both adaptation strategies, the video-based simulation responds to the source of adaptivity that was analyzed *before* learners entered Visit-Math as a learning environment. This is sometimes called meso-level adaptivity (Tetzlaff et al., 2021) or task-loop adaptivity (Aleven et al., 2017). However, micro-level adaptivity (Tetzlaff et al., 2021), for which the source of adaptivity is analyzed *while* learning, is within reach for Visit-Math. A micro-level adaptation that adapts conceptual prompts based on a real-time evaluation of the learners' notes during the assessment process was developed and is currently being investigated (Nickl, Sommerhoff, et al., 2022). For this study, the development of an adaptation strategy is currently complemented by an empirical evaluation (Plass & Pawar, 2020). As a typical approach to evaluating an adaptation strategy, researchers investigate whether the developed adaptive support is more effective than non-adaptive or no support, with early results suggesting the benefits of adaptive support over non-adaptive support (Radkowitsch et al., 2021; Sailer et al., 2022).

Altogether, the present dissertation shows a possible way of approaching adaptivity in a given (so far non-adaptive) learning environment. This approach draws on theory-based considerations but also highlights starting points for empirical evaluations, as the combination of theory and empirical evaluation is stressed as a promising approach to validly establish adaptivity in a given learning environment (Plass & Pawar, 2020).

6.3 Implications for Teacher Education

With regard to the superordinate Research Question (1), this dissertation identified three learner characteristic profiles (*knowledgeable*, *motivated*, and *potentially struggling*) that future teachers bring to the video-based simulation Visit-Math. These profiles were also shown to be related to situative learning experiences and assessment performance in the simulation. Based on these profiles, a utility value intervention and conceptual prompts were developed as promising strategies for effective scaffolding, which were then evaluated with regard to superordinate Research Question (2). The findings show that effective scaffolding strategies may consist of providing conceptual prompts or, possibly, a utility value intervention but not both together. Furthermore, the utility value intervention was most beneficial for learners with a comparably high success expectancy, whereas conceptual prompts were most beneficial for learners with a comparably low success expectancy. In this section, I briefly highlight some implications of these findings for the practice of teacher education.

As a contribution to the practice of teacher education, Visit-Math, as a scalable and always accessible online learning environment for future teachers to acquire assessment skills, was further developed (Codreanu, 2021). A utility value intervention designed based on prior promising examples (Hulleman & Harackiewicz, 2021) and conceptual prompts were implemented as promising scaffolding strategies in this improved version of the simulation, further improving the effective use of this simulation for teacher education. By identifying learner characteristic profiles within the simulation and developing scaffolding based on these profiles, the simulation can also be designed as an adaptive simulation. To further support future teachers in acquiring assessment skills, the simulation can even be adapted based on real-time measures (see Nickl, Sommerhoff, et al., 2022).

To identify the learner characteristic profiles, LPA was used. The learner characteristics that were used as indicators for LPA were related to the mathematical content (CK and PCK) and to motivational characteristics, primarily focusing assessment skills (interest, self-regulation, and self-efficacy). Thus, such profiles are likely to appear in learning settings in teacher education with regard to assessment skills in a mathematical context but may also appear in other learning settings. Thus, instructors can use the identified profiles in their own teaching practice to provide (possibly adaptive) support for learners with different learner characteristic profiles, which is expected to increase learning gains (Kulik et al., 1990; van de Pol et al., 2010). The use of LPA also showed effectiveness when approaching adaptivity *within* the Visit-Math simulation, as described in Section 6.2. In this regard, the steps described in Section 6.2 (from the choice regarding the source of adaptivity toward the development of an adaptation strategy; see Figure 12) may provide guidance for important considerations. The realization of the steps in this dissertation may also serve as a blueprint for designing adaptive learning environments within teacher education.

Another aspect that may increase learning gains, as highlighted by this dissertation, is the use of scaffolding. Primarily, based on the findings in this dissertation, this refers to scaffolding within simulations; however, this may also apply to scaffolding within other virtual learning environments or even classroom settings (Belland et al., 2017; Chernikova, Heitzmann, Fink, et al., 2020). For example, this dissertation found that a utility value intervention may lead to better learning outcomes by helping students to explore the utility of the topic to be taught by themselves, which could also be implemented in other learning contexts in teacher education (Hulleman & Harackiewicz, 2021); yet, utility value interventions have rarely been used in such contexts (Rochnia & Gräsel, 2022). To this end, implementing such scaffolding techniques in teacher education may also lead to future teachers using these strategies in their own professional practice (Bandura, 1977). When combining such motivational scaffolding with conceptual scaffolds, the present dissertation also highlights that such combinations may not necessarily be beneficial for learning, which suggests exercising caution in combining motivational and conceptual scaffolding (Cromley et al., 2020). This caution is necessary with regard to the scaffolds used in this dissertation, but it might also be necessary in other settings, maybe even classroom settings.

However, this dissertation not only highlights the importance of being cautious when combining different scaffolds but also points toward a sensitization for the interaction between scaffolds (or generally, support measures) and learners' prerequisites. To this end, the influence of the learners' success expectancy on the effectiveness of scaffolds and the preceding discussion about adaptivity substantiate that learning is an individual process, and one-size-fits-all instruction may not be the most appropriate strategy to maximize learning gains. This, again, emphasizes the advantages of individualized learning.

6.4 Limitations and Further Research

The findings of this dissertation have limitations that need to be considered when interpreting the results and conclusions. First, as a typical constraint, the sample size needs to be considered. With regard to LPA, a sample size of 150 is legitimate but in the lower range of possible sizes compared to other LPAs (Spurk et al., 2020). However, considering other LPAs conducted with (future) teachers, this sample size is average. For example, Blömeke, Hoth, et al. (2015) conducted LPA with 231 primary teachers, whereas Blömeke et al. (2020) conducted LPA with 77 secondary mathematics teachers. In Study B, the power analysis revealed that the sample size was sufficient to detect small to medium effects. Again, with regard to the context and prior studies, the obtained sample of 108 future teachers seems appropriate. However, it is desirable that further research targets larger sample sizes, which may allow for the substantiation of the results of this dissertation and may also allow the detection of smaller effects.

Second, both empirical studies were affected by COVID-19 in such a way that in-person data collections were barely possible. Thus, studies had to be conducted online, which could have led to a decrease in data quality. However, the data did not give any indication to doubt their quality. Another effect of conducting the studies online may be an increased selection effect regarding the sample: Only future teachers who already perceive a high utility of assessment skills might have participated, thus leading to a possible ceiling effect. However, with additional remuneration, this effect was probably attenuated.

Third, prior research has highlighted the domain specificity of assessment skills (Spinath, 2005). Thus, what is understood in the sense of assessment skills in this dissertation is first limited to the specific assessment situation in Visit-Math, which can be categorized as a short-term assessment for learning, which is relatively specific to certain skills of the learners (mathematical proof skills), and inaccurate judgments do not provoke negative consequences for the (simulated) student (Karst et al., 2017). Therefore, future studies need to carefully consider whether the results can be transferred to other assessment situations. This, however, is a typical constraint with regard to teachers' assessment skills.

Fourth, another limitation with regard to teachers' assessment skills is their conceptualization. As outlined in Section 2.1, the main constituents of assessment skills are dispositions, situation-specific skills, and assessment performance. This dissertation mainly focused on judgment accuracy (operationalized as the matching of the future teachers' solution with an expert's solution) as a performance measure and learning outcome. This is in line with prior research (see Urhahne & Wijnia, 2021) but may also be extended in further research. Additional measures to rate different aspects of the quality of the judgment are one possible direction for further research, such as the different measures outlined by Schrader (1989) or efficiency following future teachers' acquisition of sufficient assessment skills (Heitzmann et al., 2019). Another aspect of teachers' assessment performance is their assessment process. As the first approach, Article A draws on the number of watched videos and the number of written words in the assessment process. However, these measures are rather superficial and do not allow viable conclusions about underlying situation-specific skills and dispositions. Thus, as a next step, the notes are currently being coded, and the first set of publications, including a content-wise description (Nickl, Sommerhoff, Codreanu, et al., in press) and activity-based description (Radkowitsch et al., 2022) of the notes, are currently in the publication process. This may further allow for a description of teachers' situation-specific skills, which are also manifested in the assessment process (Blömeke et al., 2016).

Fifth, with regard to the interventions in Article B, it is important to note that both interventions were relatively short. The utility value intervention lasted about 10 minutes, which is considered to be at the lower end of the limit to effectively support learning outcomes (Rosenzweig et al., 2019). Thus, even though the utility value intervention from Article B could be expected to increase judgment accuracy, the length of the utility value intervention may have reduced its effectiveness for learning outcomes. This may also be the case with regard to conceptual prompts, which were implemented in the simulation that lasted only about 25 minutes. Again, this is in line with prior research that mainly focused on short-term interventions for assessment skills. However, it would be particularly interesting to develop and evaluate more longitudinal interventions, including different types of scaffolding and fading strategies, to foster future teachers' assessment skills.

6.5 Conclusion

Teachers' assessment skills have been shown to be an important aspect of teachers' professional practice, but they also leave room for improvement. Improvements in teachers' assessment skills can already be achieved within university teacher education through the provision of appropriate learning opportunities, which can be enriched with scaffolding to further increase learning gains. This dissertation pursued the goal of developing effective scaffolding for facilitating future mathematics teachers' assessment skills in a video-based simulation while approaching the adaptivity of the simulation. Based on empirically identified patterns of learner characteristics, promising scaffolds were derived, developed, and empirically evaluated. The dissertation makes several contributions to the field of research.

Conceptually, this dissertation identified three typical patterns of learner characteristics that future teachers may bring to a simulation and that also affect judgment accuracy. To this end, the important role of knowledge in assessment performance was substantiated, whereas the role of motivational characteristics only showed tentatively. This has led to the identification of possible directions for effective scaffolding strategies in the context of assessment skills. As possible scaffolds, conceptual prompts and a utility value intervention were designed and evaluated, showing the potential of conceptual and motivational scaffolding to increase future teachers' learning gains—as long as provided individually; a combined presentation was less effective than each scaffold individually, suggesting caution when combining conceptual and motivational scaffolding. Moreover, future teachers' success expectancy affected the effectiveness of both supporting means, underpinning the role of the learners' characteristics in the effectiveness of scaffolding, potentially indicating additional motivational effects of conceptual prompts and supporting the positive interaction of expectancy and the value component of motivation.

Methodologically, this dissertation presented a way to design scaffolding based on theoretical considerations and an empirical evaluation of what learners bring to the video-based simulation. This approach highlighted advantages of LPA as a person-centered approach for developing effective scaffolding compared to variable-centered approaches, as it allowed to identify patterns that could be used to analyze pattern by pattern, what possible scaffolding strategies could be. By using a 2 × 2 intervention design for investigating the effectiveness of the two derived scaffolds, it was possible not only to evaluate each individual scaffold's effectiveness but also to identify the combination of both scaffolds as less promising. Altogether, this approach to designing effective scaffolding was also interpreted in the sense of adaptivity following a step-wise approach to establishing adaptivity in learning environments that may serve as a blueprint for designing further adaptive learning environments.

With regard to practice, this dissertation designed scaffolding to facilitate future teachers' assessment skills in a given learning environment. To this end, it highlighted the role of learner characteristics as prerequisites for learning and substantiated the use of scaffolding to support learners. Lastly, this dissertation highlighted the role of scaffolding in computer-based settings beyond one-size-fits-all instructional support toward individualized support measures.

7 References

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*(3), 545–561. https://doi.org/10.1037/0022-0663.94.3.545

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer & P. A. Alexander (Eds.), *Educational psychology handbook series. Handbook of research on learning and instruction* (pp. 522–560). Routledge.

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*(1), 87–112. https://doi.org/10.1007/s12186-018-9211-5

Arnold, K.-H., Gröschner, A., & Hascher, T. (Eds.). (2014). *Schulpraktika in der Lehrerbildung/Pedagogical field experiences in teacher education: Theoretische Grundlagen, Konzeptionen, Prozesse und Effekte/Theoretical foundations, programmes, processes, and effects*. Waxmann. https://elibrary.utb.de/doi/book/10.31244/9783830980575

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

Aschbacher, P., & Alonzo, A. (2006). Examining the utility of elementary science notebooks for formative assessment purposes. *Educational Assessment*, *11*(3-4), 179–203. https://doi.org/10.1080/10627197.2006.9652989

Bandura, A. (1977). *Social learning theory*. Prentice Hall.

Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 139–145. https://doi.org/10.1024/1010-0652.23.2.139

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, *40*(1), 193–211. https://doi.org/10.1007/s11251-011-9167-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Professional competencies of teachers]. *Zeitschrift Für Erziehungswissenschaft*, *9*(4), 469–520. https://doi.org/10.1007/s11618-006-0165-2

Belland, B. R. (2011). Distributed cognition as a lens to understand the effects of scaffolding: The role of transfer of responsibility. *Educational Psychology Review*, *23*(4), 577–600. https://doi.org/10.1007/s10648-011-9176-5

Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A framework for designing scaffoldings that improve motivation and cognition. *Educational Psychologist*, *48*(4), 243–270. https://doi.org/10.1080/00461520.2013.838920

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

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, *17*(5), 564–577. https://doi.org/10.1016/j.learninstruc.2007.09.007

Betz, A., Flake, S., Mierwald, M., & Vanderbeke, M. (2016). Modelling authenticity in teaching and learning contexts: A contribution to theory development and empirical investigation of the construct. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016.* International Society of the Learning Sciences.

Bixler, B. A. (2007). *The effects of scaffolding student's problem-solving process via question prompts on problem solving and intrinsic motivation in an online learning* *environment* [Doctoral dissertation, Pennsylvania State University]. ProQuest Dissertations Publishing. https://etda.libraries.psu.edu/files/final_submissions/838

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

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

Blömeke, S., Hoth, J., Döhrmann, M., Busse, A., Kaiser, G., & König, J. (2015). Teacher change during induction: Development of beginning primary teachers' knowledge, beliefs, and performance. *International Journal of Science and Mathematics Education*, *13*(2), 287–308. https://doi.org/10.1007/s10763-015-9619-4

Blömeke, S., & Kaiser, G. (2017). Understanding the development of teachers' professional competencies as personally, situationally, and socially determined. In D. J. Clandinin & J. Husu (Eds.), *The SAGE handbook of research on teacher education* (pp. 783–802). SAGE Publications.

Blömeke, S., Kaiser, G., König, J., & Jentsch, A. (2020). Profiles of mathematics teachers' competence and their relation to instructional quality. *ZDM*, *52*(2), 329–342. https://doi.org/10.1007/s11858-020-01128-y

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3–27. http://www.jstor.org/stable/25791751

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, *24*(2), 417–436. https://doi.org/10.1016/j.tate.2006.11.012

Brandl, L., Richters, C., Radkowitsch, A., Obersteiner, A., & Stadler, M. (2021). Simulation-based learning of complex skills: Predicting performance with theoretically derived process features. *Psychological Test and Assessment Modeling*, *63*(4), 542–560.

Braun, L. T., Zottmann, J. M., Adolf, C., Lottspeich, C., Then, C., Wirth, S., Fischer, M. R., & Schmidmaier, R. (2017). Representation scaffolds improve diagnostic efficiency in medical students. *Medical Education*, *51*(11), 1118–1126. https://doi.org/10.1111/medu.13355

Brophy, J. E. (Ed.). (2004). *Advances in research on teaching: Vol. 10. Using video in teacher education*. Emerald Group Publishing.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1982, June). *Learning, remembering, and understanding.* (Technical Report No. 244). University of Illinois at Urbana-Champaign, Center for the Study of Reading. https://eric.ed.gov/?id=ED217401

Brunner, M., Anders, Y., Hachfeld, A., & Krauss, S. (2013). The diagnostic skills of mathematics teachers. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 229–248). Springer. https://doi.org/10.1007/978-1-4614-5149-5_11

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217. https://doi.org/10.1037/h0047470

Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *Lecture Notes in Computer Science. The Adaptive Web* (Vol. 4321, pp. 3–53). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_1

Canning, E. A., & Harackiewicz, J. M. (2015). Teach it, don't preach it: The differential effects of directly-communicated and self-generated utility value information. *Motivation Science*, *1*(1), 47–71. https://doi.org/10.1037/mot0000015

Chen, C.-H., & Wu, I.-C. (2012). The interplay between cognitive and motivational variables in a supportive online learning system for secondary physical education. *Computers & Education*, *58*(1), 542–550. https://doi.org/10.1016/j.compedu.2011.09.012

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2020). Facilitating diagnostic competences in higher education—A meta-analysis in medical and teacher education. *Educational Psychology Review*, *32*(1), 157–196. https://doi.org/10.1007/s10648-019-09492-2

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

Chernikova, O., Holzberger, D., Heitzmann, N., Stadler, M., Seidel, T., & Fischer, F. (in press). Where salience goes beyond authenticity: A meta-analysis on simulation-based learning in higher education. *Zeitschrift Für Pädagogische Psychologie*.

Codreanu, E. (2021). *Approximating practice: Developing a video-based simulation for measuring preservice teachers' diagnostic skills* [Doctoral dissertation, Technical University of Munich]. mediaTUM. http://mediatum.ub.tum.de.eaccess.ub.tum.de/?id=1609556

Codreanu, E., Huber, S., Reinhold, S., Sommerhoff, D., Neuhaus, B. J., Schmidmaier, R., Ufer, S., & Seidel, T. (2022). Diagnosing mathematical argumentation skills: A video-based simulation for pre-service teachers. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations: Examples from teacher education and medical education* (pp. 33–47). Springer International Publishing. https://doi.org/10.1007/978-3-030-89147-3_4

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

Codreanu, E., Sommerhoff, D., Huber, S., Ufer, S., & Seidel, T. (2021). Exploring the process of preservice teachers' diagnostic activities in a video-based simulation. *Frontiers in Education*, *6*, Article 626666. https://doi.org/10.3389/feduc.2021.626666

Cohen, E., Hoz, R., & Kaplan, H. (2013). The practicum in preservice teacher education: A review of empirical studies. *Teaching Education*, *24*(4), 345–380. https://doi.org/10.1080/10476210.2012.711815

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Lawrence Erlbaum Associates.

Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J., & Hatala, R. (2013). Technology-enhanced simulation to assess health professionals: A systematic review of validity evidence, research methods, and reporting quality. *Academic Medicine*, *88*(6), 872–883. https://doi.org/10.1097/ACM.0b013e31828ffdcf

COSIMA Research Unit (2021). *COSIMA framework* [Graphic]. https://www.en.for2385.uni-muenchen.de/cosima_framemodel1/cosima_frame-model_short_eng.pdf

Council of Chief State School Officers. (2013). *Interstate teacher assessment and support consortium inTASC model core teaching standards and learning progressions for teachers 1.0: A resource for ongoing teacher development.* Council of Chief State School Officers.

Cromley, J. G., Perez, T., Kaplan, A., Dai, T., Mara, K., & Balsai, M. J. (2020). Combined cognitive–motivational modules delivered via an LMS increase undergraduate biology grades. *Technology, Mind, and Behavior*, *1*(2). https://doi.org/10.1037/tmb0000020

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, *28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Dicke, T., Holzberger, D., Kunina-Habenicht, O., Linninger, C., & Schulze-Stocker, F. (2016). „Doppelter Praxisschock“ auf dem Weg ins Lehramt? Verlauf und potenzielle Einflussfaktoren emotionaler Erschöpfung während des Vorbereitungsdienstes und nach dem Berufseintritt [Double practice shock on the way to being a teacher? Course and potentially influencing factors on emotional exhaustion during pre-service and after entering the profession]. *Psychologie in Erziehung Und Unterricht*, *63*(4), 244. https://doi.org/10.2378/peu2016.art20d

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research*, *50*(4), 241–250. https://doi.org/10.1016/j.ijer.2011.08.005

Durik, A. M., Shechter, O. G., Noh, M., Rozek, C. S., & Harackiewicz, J. M. (2015). What if I can't? Success expectancies moderate the effects of utility value information on situational interest and performance. *Motivation and Emotion*, *39*(1), 104–118. https://doi.org/10.1007/s11031-014-9419-0

Dziak, J. J., Bray, B. C., Zhang, J., Zhang, M., & Lanza, S. T. (2016). Comparing the performance of improved classify-analyze approaches for distal outcomes in latent profile analysis. *Methodology*, *12*(4), 107–116. https://doi.org/10.1027/1614-2241/a000114

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Estapa, A., & Amador, J. M. (2023). A qualitative metasynthesis of video-based prompts and noticing in mathematics education. *Mathematics Education Research Journal*, *35*(1), 105–131. https://doi.org/10.1007/s13394-021-00378-7

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*(5), 797–813. https://doi.org/10.1037/0022-3514.91.5.797

Fink, M. C., Reitmeier, V., Stadler, M., Siebeck, M., Fischer, F., & Fischer, M. R. (2021). Assessment of diagnostic competences with standardized patients versus virtual patients: Experimental study in the context of history taking. *Journal of Medical Internet Research*, *23*(3), e21196. https://doi.org/10.2196/21196

Fischer, F., Bauer, E., Seidel, T., Schmidmaier, R., Radkowitsch, A., Neuhaus, B. J., Hofer, S. I., Sommerhoff, D., Ufer, S., Kuhn, J., Küchemann, S., Sailer, M., Koenen, J., Gartmeier, M., Berberat, P., Frenzel, A., Heitzmann, N., Holzberger, D., Pfeffer, J., . . . Fischer, M. R. (2022). Representational scaffolding in digital simulations – Learning professional practices in higher education. *Information and Learning Sciences*, *123*(11/12), 645–665. https://doi.org/10.1108/ILS-06-2022-0076

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

Fischer, F., & Opitz, A. (Eds.). (2022). *Learning to diagnose with simulations: Examples from teacher education and medical education.* Springer International Publishing.

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M., Girwidz, R., Obersteiner, A., Reiss, K., Stürmer, K., Siebeck, M., Schmidmaier, R., Seidel, T., Ufer, S., Wecker, C., & Neuhaus, B. (2018). Systematizing professional knowledge of medical doctors and teachers: Development of an interdisciplinary framework in the context of diagnostic competences. *Education Sciences*, *8*(4), 207. https://doi.org/10.3390/educsci8040207

Frank, B. (2015). *Presence messen in laborbasierter Forschung mit Mikrowelten: Entwicklung und erste Validierung eines Fragebogens zur Messung von Presence* [Measuring presence in laboratory-based research with microworlds: development and first validation of a presence questionnaire]. *BestMasters*. Springer Wiesbaden. https://doi.org/10.1007/978-3-658-08148-5

Friedrich, H. F., & Mandl, H. (1992). Lern- und Denkstrategien - ein Problemaufriß. In H. Mandl & H. F. Friedrich (Eds.), *Lern- und Denkstrategien: Analyse und Intervention* (pp. 1–54). Hogrefe.

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. https://doi.org/10.1037/dev0000028

Gaspard, H., Parrisius, C., Piesch, H., Kleinhansl, M., Wille, E., Nagengast, B., Trautwein, U., & Hulleman, C. S. (2021). The potential of relevance interventions for scaling up: A cluster-randomized trial testing the effectiveness of a relevance intervention in math classrooms. *Journal of Educational Psychology*, *113*(8), 1507–1528. https://doi.org/10.1037/edu0000663

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

Gilbert, S. B. (2016). Perceived realism of virtual environments depends on authenticity. *Presence: Teleoperators and Virtual Environments*, *25*(4), 322–324. https://doi.org/10.1162/PRES_a_00276

Gold, B., Förster, S., & Holodynski, M. (2013). Evaluation eines videobasierten Trainingsseminars zur Förderung der professionellen Wahrnehmung von Klassenführung im Grundschulunterricht* [Evaluation of a video-based training to foster the professional vision of classroom management in elementary classrooms]. *Zeitschrift Für Pädagogische Psychologie*, *27*(3), 141–155. https://doi.org/10.1024/1010-0652/a000100

Grossman, P. (2018). *Teaching core practices in teacher education*. Harvard Education Press.

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.

Hallquist, M. N., & Wiley, J. F. (2018). Mplusautomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 621–638. https://doi.org/10.1080/10705511.2017.1402334

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

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology*, *111*(5), 745–765. https://doi.org/10.1037/pspp0000075

Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online*, *11*(2), 169–191. https://doi.org/10.25656/01:18004

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://www.taylorfrancis.com/books/mono/10.4324/9780203887332/visible-learning-john-hattie https://doi.org/10.4324/9780203887332

Heckhausen, H. (1967). *The anatomy of achievement motivation*. Academic Press.

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

Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching and Teacher Education*, *3*(2), 91–98. https://doi.org/10.1016/0742-051X(87)90010-2

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

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Holzberger, D., Maurer, C., Kunina-Habenicht, O., & Kunter, M. (2021). Ready to teach? A profile analysis of cognitive and motivational-affective teacher characteristics at the end of pre-service teacher education and the long-term effects on occupational well-being. *Teaching and Teacher Education*, *100*, 103285. https://doi.org/10.1016/j.tate.2021.103285

Horz, H., Winter, C., & Fries, S. (2009). Differential benefits of situated instructional prompts. *Computers in Human Behavior*, *25*(4), 818–828. https://doi.org/10.1016/j.chb.2008.07.001

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

Huk, T., & Ludwigs, S. (2009). Combining cognitive and affective support in order to promote learning. *Learning and Instruction*, *19*(6), 495–505. https://doi.org/10.1016/j.learninstruc.2008.09.001

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*(4), 880–895. https://doi.org/10.1037/a0019506

Hulleman, C. S., & Harackiewicz, J. M. (2021). The utility-value intervention. In G. M. Walton & A. J. Crum (Eds.), *Handbook of wise interventions: How social psychology can help people change* (pp. 100–125). The Guilford Press.

Ingenkamp, K.-H., & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* [Textbook on educational assessment]. Beltz. http://www.content-select.com/index.php?id=bib_view&ean=9783407291417

Irmer, M., Traub, D., Kramer, M., Förtsch, C., & Neuhaus, B. J. (2022). Scaffolding pre-service biology teachers’ diagnostic competences in a video-based learning environment: Measuring the effect of different types of scaffolds. *International Journal of Science Education*, 1–21. https://doi.org/10.1080/09500693.2022.2083253

Kabassi, K., & Virvou, M. (2004). Personalised adult e-training on computer use based on multiple attribute decision making. *Interacting with Computers*, *16*(1), 115–132. https://doi.org/10.1016/j.intcom.2003.11.006

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *The American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kane, M. T., & Wools, S. (2020). Perspectives on the validity of classroom assessments. In S. M. Brookhart & J. H. McMillan (Eds.), *NCME applications of educational measurement and assessment book series. Classroom assessment and educational measurement* (pp. 11–26). Routledge. https://doi.org/10.4324/9780429507533-2

Karst, K., Klug, J., & Ufer, S. (2017). Strukturierung diagnostischer Situationen im inner- und außerunterrichtlichen Handeln von Lehrkräften [Structuring assessment situations for teachers inside and outside the classroom]. In A. Südkamp & A.-K. Praetorius (Eds.), *Pädagogische Psychologie und Entwicklungspsychologie: Vol. 94. Diagnostische Kompetenz von Lehrkräften: Theoretische und methodische Weiterentwicklungen* (pp. 102–114). Waxmann.

Karst, K., Schoreit, E., & Lipowsky, F. (2014). Diagnostische Kompetenzen von Mathematiklehrern und ihr Vorhersagewert für die Lernentwicklung von Grundschulkindern [Diagnostic competencies of math teachers and their impact on learning development of elementary school children]. *Zeitschrift Für Pädagogische Psychologie*, *28*(4), 237–248. https://doi.org/10.1024/1010-0652/a000133

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2021). Making an effort versus experiencing load. *Frontiers in Education*, *6*, Article 645284, 56. https://doi.org/10.3389/feduc.2021.645284

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

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of selfregulatory patterns. *Journal of Educational Psychology*, *100*(3), 702–715. https://doi.org/10.1037/0022-0663.100.3.702

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, *6*(1), 3–34. https://doi.org/10.1007/BF00155578

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

Kosel, C., Holzberger, D., & Seidel, T. (2021). Identifying expert and novice visual scanpath patterns and their relationship to assessing learning-relevant student characteristics. *Frontiers in Education*, *5*, Article 612175, 284. https://doi.org/10.3389/feduc.2020.612175

Kosel, C., Wolter, I., & Seidel, T. (2021). Profiling secondary school students in mathematics and German language arts using learning-relevant cognitive and motivational-affective characteristics. *Learning and Instruction*, *73*, 101434. https://doi.org/10.1016/j.learninstruc.2020.101434

Kramer, M., Förtsch, C., Boone, W. J., Seidel, T., & Neuhaus, B. J. (2021). Investigating pre-service biology teachers' diagnostic competences: Relationships between professional knowledge, diagnostic activities, and diagnostic accuracy. *Education Sciences*, *11*(3), 89. https://doi.org/10.3390/educsci11030089

Kramer, M., Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation*, *68*, 100973. https://doi.org/10.1016/j.stueduc.2020.100973

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

Kron, S., Sommerhoff, D., Achtner, M., Stürmer, K., Wecker, C., Siebeck, M., & Ufer, S. (2022). Cognitive and motivational person characteristics as predictors of diagnostic performance: Combined effects on pre-service teachers’ diagnostic task selection and accuracy. *Journal Für Mathematik-Didaktik*, *43*(1), 135–172. https://doi.org/10.1007/s13138-022-00200-2

Kron, S., Sommerhoff, D., Achtner, M., & Ufer, S. (2021). Selecting mathematical tasks for assessing student’s understanding: Pre-service teachers’ sensitivity to and adaptive use of diagnostic task potential in simulated diagnostic one-to-one interviews. *Frontiers in Education*, *6*, Article 604568. https://doi.org/10.3389/feduc.2021.604568

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, *60*(2), 265–299. https://doi.org/10.3102/00346543060002265

Kultusministerkonferenz. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften: Beschluss der Kultusministerkonferenz vom 16.12.2004 i. d. F. vom 16.05.2019* [Standards for Teacher Education: Educational Sciences: Resolution of Standing Conference of the Ministers of Education and Cultural Affairs of Germany from December 16, 2004, in the version dated May 16, 2019]. Berlin.

Kunter, M., Baumert, J., Leutner, D., Terhart, E., Seidel, T., Dicke, T., Holzberger, D., Kunina-Habenicht, O., Linninger, C., Lohse-Bossenz, H., Schulze-Stocker, F., & Stürmer, K. (2016). *Dokumentation der Erhebungsinstrumente der Projektphasen des BilWiss-Forschungsprogramms von 2009 bis 2016: Bildungswissenschaftliches Wissen und der Erwerb professioneller Kompetenz in der Lehramtsausbildung (BilWiss); die Bedeutung des bildungswissenschaftlichen Hochschulwissens für den Berufseinstieg von Lehrkräften (BilWiss-Beruf)* [Scale manual of BilWiss research program from 2009 to 2016]. Goethe-Universität.

Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2013). The development of teachers’ professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 63–77). Springer. https://doi.org/10.1007/978-1-4614-5149-5_4

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

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente* [PISA 2000: Scale manual]. *Materialien aus der Bildungsforschung: Nr. 72*. Max-Planck-Inst. für Bildungsforschung. http://hdl.handle.net/hdl:11858/00-001M-0000-0023-9987-C

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Houghton Mifflin.

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Leiss, D. (2010). Adaptive Lehrerinterventionen beim mathematischen Modellieren – empirische Befunde einer vergleichenden Labor- und Unterrichtsstudie [Adaptive teacher interventions in mathematical modelling—empirical findings of a comparative laboratory and classroom study]. *Journal für Mathematik-Didaktik*, *31*(2), 197–226. https://doi.org/10.1007/s13138-010-0013-z

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means. R package* (version 1.7.5.) [Computer software]. https://CRAN.R-project.org/package=emmeans

Levin, D. M., Hammer, D., & Coffey, J. E. (2009). Novice teachers' attention to student thinking. *Journal of Teacher Education*, *60*(2), 142–154. https://doi.org/10.1177/0022487108330245

Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' diagnostic judgments by cognitive modeling (DiaCoM). *Teaching and Teacher Education*, *91*, 103059. https://doi.org/10.1016/j.tate.2020.103059

Mamede, S., & Schmidt, H. G. (2017). Reflection in medical diagnosis: A literature review. *Health Professions Education*, *3*(1), 15–25. https://doi.org/10.1016/j.hpe.2017.01.003

Mayer, R. E. (Ed.). (2014). *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 131–152). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.009

McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern [Teachers' diagnostic skills to judge student performance and task difficulty when learning materials include instructional images]. *Zeitschrift Für Pädagogische Psychologie*, *23*(34), 223–235. https://doi.org/10.1024/1010-0652.23.34.223

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

Moreno, R., & Mayer, R. E. (2010). Techniques that increase generative processing in multimedia learning: Open questions for cognitive load research. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 153–178). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.010

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, *22*(8), 1058–1066. https://doi.org/10.1177/0956797611415540

Nakic, J., Granic, A., & Glavinic, V. (2015). Anatomy of student models in adaptive learning systems: A systematic literature review of individual differences from 2001 to 2013.

Journal of Educational Computing Research, *51*(4), 459–489. https://doi.org/10.2190/EC.51.4.e

Nickl, M., Sommerhoff, D., Radkowitsch, A., Huber, S. A., Ufer, S., Plass, J. L., & Seidel, T. (2022). *Effects of real-time adaptivity in facilitating assessment skills of mathematics pre-service teachers through scaffolding in simulations.* https://doi.org/10.17605/OSF.IO/GK58D

Nickl, M., Sommerhoff, D., Codreanu, E., Ufer, S., & Seidel, T. (in press). The Role of Teachers' Person Characteristics for Assessing Students' Proof Skills. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education.* PME.

Nückles, M., Wittwer, J., & Renkl, A. (2005). Information about a layperson's knowledge supports experts in giving effective and efficient online advice to laypersons. *Journal of Experimental Psychology. Applied*, *11*(4), 219–236. https://doi.org/10.1037/1076-898X.11.4.219

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

Oudman, S., van de Pol, J., Bakker, A., Moerbeek, M., & van Gog, T. (2018). Effects of different cue types on the accuracy of primary school teachers' judgments of students' mathematical understanding. *Teaching and Teacher Education*, *76*, 214–226. https://doi.org/10.1016/j.tate.2018.02.007

Palincsar, A. S. (1998). Keeping the metaphor of scaffolding fresh—A response to C. Addison Stone's "The metaphor of scaffolding: Its utility for the field of learning disabilities". *Journal of Learning Disabilities*, *31*(4), 370–373. https://doi.org/10.1177/002221949803100406

Petko, D. (2014). *Einführung in die Mediendidaktik: Lehren und Lernen mit digitalen Medien* [Introduction to media education]. *Bildungswissen Lehramt* (Vol. 25). Beltz.

Pickal, A. J., Wecker, C., Neuhaus, B., & Girwidz, R. (2022). Learning to diagnose secondary school students' scientific reasoning skills in physics and biology: Video-based simulations for pre-service teachers. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations: Examples from teacher education and medical education.* Springer International Publishing.

Pieger, E., & Bannert, M. (2018). Differential effects of students' self-directed metacognitive prompts. *Computers in Human Behavior*, *86*, 165–173. https://doi.org/10.1016/j.chb.2018.04.022

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of selfregulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3

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

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*(1), 64–76. https://doi.org/10.1080/00220671.2012.667010

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? [Do students with optimistic self-esteem profit?]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *48*(1), 14–26. https://doi.org/10.1026/0049-8637/a000140

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, *13*(3), 337–386. http://www.jstor.org/stable/1466941

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Radkowitsch, A., Fischer, M. R., Schmidmaier, R., & Fischer, F. (2020). Learning to diagnose collaboratively: Validating a simulation for medical students. *GMS Journal for Medical Education*, *37*(5), Doc51. https://doi.org/10.3205/zma001344

Radkowitsch, A., Sailer, M., Schmidmaier, R., Fischer, M. R., & Fischer, F. (2021). Learning to diagnose collaboratively – Effects of adaptive collaboration scripts in agent-based medical simulations. *Learning and Instruction*, *75*, 101487. https://doi.org/10.1016/j.learninstruc.2021.101487

Radkowitsch, A., Sommerhoff, D., Nickl, M., Codreanu, E., Ufer, S., & Seidel, T. (2022). Exploring the diagnostic process of pre-service teachers using a simulation – A latent profile approach. In C. Chinn, E. Tan, C. Chan, & Y. Kali (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences - ICLS 2022* (pp. 567–574). International Society of the Learning Sciences.

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen [FAM: A questionnaire for measuring situative motivation in learning and performance situations]. *Diagnostica*, *47*(2), 57–66. https://doi.org/10.1026//0012-1924.47.2.57

Riese, J., & Reinhold, P. (2009). *Measuring physics student teachers' pedagogical content knowledge as an indicator of their professional action competence.* European Science Education Research Association, Istanbul, Turkey.

Rieu, A., Leuders, T., & Loibl, K. (2022). Teachers' diagnostic judgments on tasks as information processing – The role of pedagogical content knowledge for task diagnosis. *Teaching and Teacher Education*, *111*, 103621. https://doi.org/10.1016/j.tate.2021.103621

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*(2), 340–352. https://doi.org/10.1037/0022-3514.80.2.340

Rochnia, M., & Gräsel, C. (2022). Can the utility value of educational sciences be induced based on a reflection example or empirical findings—Or just somehow? *Frontiers in Education*, *7*, Article 1006079, 979. https://doi.org/10.3389/feduc.2022.1006079

Rosenzweig, E. Q., Hulleman, C. S., Barron, K. E., Kosovich, J. J., Priniski, S. J., & Wigfield, A. (2019). Promises and pitfalls of adapting utility value interventions for online math courses. *The Journal of Experimental Education*, *87*(2), 332–352. https://doi.org/10.1080/00220973.2018.1496059

Rosenzweig, E. Q., Wigfield, A., & Eccles, J. S. (2022). Beyond utility value interventions: The why, when, and how for next steps in expectancy-value intervention research. *Educational Psychologist*, *57*(1), 11–30. https://doi.org/10.1080/00461520.2021.1984242

Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *Journal of Educational Psychology*, *112*(1), 166–182. https://doi.org/10.1037/edu0000370

Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: Thirst for knowledge. *Learning and Instruction*, *32*, 37–50. https://doi.org/10.1016/j.learninstruc.2014.01.002

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2022). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learning and Instruction*, 101620. https://doi.org/10.1016/j.learninstruc.2022.101620

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

Schiefele, U., Streblow, L., & Retelsdorf, J. (2013). Dimensions of teacher interest and their relations to occupational well-being and instructional practices. *Journal for Educational Research Online*, *5*(1), 7–37.

Schons, C., Obersteiner, A., Reinhold, F., Fischer, F., & Reiss, K. (2022). Developing a simulation to foster prospective mathematics teachers' diagnostic competencies: The effects of scaffolding. *Journal Für Mathematik-Didaktik*, 1–24. https://doi.org/10.1007/s13138-022-00210-0

Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts* [Teachers' assessment skills and their relevance to the design and effectiveness of teaching]. *Europäische Hochschulschriften. Reihe 6: Psychologie* (Vol. 289). Peter Lang.

Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen [Teacher diagnosis and diagnostic competence]. *Beiträge zur Lehrerbildung*, *31*(2), 154–165. https://www.pedocs.de/frontdoor.php?source_opus=13843

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education*, *34*, 56–65. https://doi.org/10.1016/j.tate.2013.03.004

Seidel, T., Blomberg, G., & Stürmer, K. (2010). „Observer" – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. Projekt OBSERVE ["Observer" - Validating a video-based instrument for measuring professional vision. Project OBSERVE]. *Zeitschrift für Pädagogik, Beiheft, 56*, 296–306.

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

Seidel, T., Stürmer, K., 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

Selzer, M. N., Gazcon, N. F., & Larrea, M. L. (2019). Effects of virtual presence and learning outcome using low-end virtual reality systems. *Displays*, *59*, 9–15. https://doi.org/10.1016/j.displa.2019.04.002

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4. https://doi.org/10.2307/1175860

Shute, V. J. (1995). Smart: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Interaction*, *5*(1), 1–44. https://doi.org/10.1007/BF01101800

Shute, V. J., & Zapata-River, D. (2012). Adaptive educational systems. In P. J. Durlach & A. M. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 7–27). Cambridge University Press.

Sommerhoff, D. (2017). *The individual cognitive resources underlying students' mathematical argumentation and proof skills* [Doctoral dissertation, Ludwig-Maximilians-Universität München]. University Library of LMU Munich. https://edoc.ub.uni-muenchen.de/22687

Sommerhoff, D., Codreanu, E., Nickl, M., Ufer, S., & Seidel, T. (2023). Pre-service teachers' learning of diagnostic skills in a video-based simulation: Effects of conceptual vs. interconnecting prompts on judgment accuracy and the diagnostic process. *Learning and Instruction*, 101689. https://doi.org/10.1016/j.learninstruc.2022.101689

Song, J., & Chung, Y. (2020). Reexamining the interaction between expectancy and task value in academic settings. *Learning and Individual Differences*, *78*, 101839. https://doi.org/10.1016/j.lindif.2020.101839

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teacher judgments on student characteristics and the construct of diagnostic competence]. *Zeitschrift Für Pädagogische Psychologie*, *19*(1/2), 85–95. https://doi.org/10.1024/1010-0652.19.12.85

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *The School Review*, *20*(7), 442–457. http://www.jstor.org/stable/1076706

Starch, D., & Elliott, E. C. (1913). Reliability of grading work in mathematics. *The School Review*, *21*(4), 254–259. http://www.jstor.org/stable/1076246

Stevens, J. A., & Kincaid, J. P. (2015). The relationship between presence and performance in virtual simulation training. *Open Journal of Modelling and Simulation*, *03*(02), 41–48. https://doi.org/10.4236/ojmsi.2015.32005

Stokking, K., Leenders, F., Jong, J. de, & van Tartwijk, J. (2003). From student to teacher: Reducing practice shock and early dropout in the teaching profession. *European Journal of Teacher Education*, *26*(3), 329–350. https://doi.org/10.1080/0261976032000128175

Stone, C. A. (1998). The metaphor of scaffolding: Its utility for the field of learning disabilities. *Journal of Learning Disabilities*, *31*(4), 344–364. https://doi.org/10.1177/002221949803100404

Stürmer, K., Marczynski, B., Wecker, C., Siebeck, M., & Ufer, S. (2021). Praxisnahe Lerngelegenheiten in der Lehrerbildung - Validierung der simulationsbasierten Lernumgebung DiMaL zur Förderung diagnostischer Kompetenzen von angehenden Mathematiklehrpersonen [Practical learning opportunities in teacher education - Validation of the simulation-based learning environment DiMaL for the promotion of diagnostic competences of prospective mathematics teachers]. In N. Beck, T. Bohl, & S. Meissner (Eds.), *Vielfältig herausgefordert: Forschungs- und Entwicklungsfelder der Lehrerbildung auf dem Prüfstand* (pp. 57–72). Tübingen University Press.

Stürmer, K., Seidel, T., & Schäfer, S. (2013). Changes in professional vision in the context of practice. *Gruppendynamik Und Organisationsberatung*, *44*(3), 339–355. https://doi.org/10.1007/s11612-013-0216-0

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. https://doi.org/10.1037/a0027627

Südkamp, A., Möller, J., & Pohlmann, B. (2008). Der Simulierte Klassenraum: Eine experimentelle Untersuchung zur diagnostischen Kompetenz [The simulated classroom: An experimental study on diagnostic competence]. *Zeitschrift Für Pädagogische Psychologie*, *22*(34), 261–276. https://doi.org/10.1024/1010-0652.22.34.261

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer New York, NY. https://doi.org/10.1007/978-1-4419-8126-4_5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

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

Tetzlaff, L., Schmiedek, F., & Brod, G. (2021). Developing personalized education: A dynamic framework. *Educational Psychology Review*, *33*(3), 863–882. https://doi.org/10.1007/s10648-020-09570-w

Todorova, M., Sunder, C., Steffensky, M., & Möller, K. (2017). Pre-service teachers' professional vision of instructional support in primary science classes: How content-specific is this skill and which learning opportunities in initial teacher education are relevant for its acquisition? *Teaching and Teacher Education*, *68*, 275–288. https://doi.org/10.1016/j.tate.2017.08.016

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*(3), 763–777. https://doi.org/10.1037/a0027470

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*(2), 202–248. https://doi.org/10.3102/00346543068002202

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. http://www.jstor.org/stable/1738360

Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, *32*, 100374. https://doi.org/10.1016/j.edurev.2020.100374

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

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. https://www.learntechlib.org/primary/p/9171/

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130. https://doi.org/10.1016/j.chb.2010.07.038

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. https://doi.org/10.1093/pan/mpq025

Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, F., Böcking, S., Schramm, H., Gysbers, A., Hartmann, T., Klimmt, C., Laarni, J., Ravaja, N., Sacau, A., Baumgartner, T., & Jäncke, P. (2004). MEC spatial presence questionnaire (MEC-SPQ): Short documentation and instructions for application. *Report to the European Community, Project Presence: MEC (IST-2001-37661)*. http://www.ijk.hmt-hannover.de/presence

Voss, T. (2022). Not useful to inform teaching practice? Student teachers hold skeptical beliefs about evidence from education science. *Frontiers in Education*, *7*, Article 976791. https://doi.org/10.3389/feduc.2022.976791

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Weresch-Deperrois, I., Bodensohn, R., & Jäger, R. S. (2009). Curriculare Standards in der Praxis: Einschätzung ihres Stellenwerts, ihrer Anwendungshäufigkeit, Schwierigkeit und Bedeutung in der Lehrerausbildung und universitären Vorbereitung im Bachelor-Studium der Lehrerbildung. Eine Erkundungsstudie [Curricular standards in practice: Appraisement of their importance, frequency of application, difficulty, relevance in teacher education and preparation by the university in the field of BA-teacher education. An explorative study]. *Lehrerbildung auf dem Prüfstand*, *2*(2), 324–345. https://doi.org/10.25656/01:15852

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*(1), 49–78. https://doi.org/10.1007/BF02209024

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Eccles, J. S. (2020). Chapter Five - 35 years of research on students' subjective task values and motivation: A look back and a look forward. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 7, pp. 161–198). Elsevier. https://doi.org/10.1016/bs.adms.2019.05.002

Wild, K.-P., & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens [Learning strategies of university students: Factor structure and reliability of a new questionnaire]. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *15*(4), 185–200.

Wildgans-Lang, A., Scheuerer, S., Obersteiner, A., Fischer, F., & Reiss, K. (2020). Analyzing prospective mathematics teachers' diagnostic processes in a simulated environment. *ZDM*, *52*(2), 241–254. https://doi.org/10.1007/s11858-020-01139-9

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Woo, S. E., Jebb, A. T., Tay, L., & Parrigon, S. (2018). Putting the "Person" in the Center. *Organizational Research Methods*, *21*(4), 814–845. https://doi.org/10.1177/1094428117752467

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

Zumbach, J., Rammerstorfer, L., & Deibl, I. (2020). Cognitive and metacognitive support in learning with a serious game about demographic change. *Computers in Human Behavior*, *103*, 120–129. https://doi.org/10.1016/j.chb.2019.09.026

Appendix

Article A

Nickl, M., Huber, S. A., Sommerhoff, D., Codreanu, E., Ufer, S., & Seidel, T. (2022). Video-based simulations in teacher education: The role of learner characteristics as capacities for positive learning experiences and high performance. *International Journal of Educational Technology in Higher Education*, *19*, Article 45. https://doi.org/10.1186/s41239-022-00351-9

Article B

Nickl, M., Sommerhoff, D., Böheim, R., Ufer, S., & Seidel, T. (2023). Fostering pre-service teachers' assessment skills in a video simulation: Differential effects of a utility value intervention and conceptual knowledge prompts. *Zeitschrift für Pädagogische Psychologie*. https://doi.org/10.1024/1010-0652/a000362

Note: