Indicators of Diagnostic Competence

—

Investigating Pre-Service Teachers' Resources, Diagnostic Processes, and Diagnostic Performance using Simulations

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig-Maximilians-Universität München

eingereicht von

Stephanie Kron

13.12.2023

1. Gutachter: Prof. Dr. Stefan Ufer
2. Gutachter: Prof. Dr. Andreas Obersteiner
3. Gutachter: Prof. Dr. Knut Neumann

Tag der Disputation: 11.04.2024

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5.)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Kron, Stephanie
Name, Vorname

München, 13.12.2023
Ort, Datum

Stephanie Kron
Unterschrift Doktorand/in

Acknowledgements

“The most valuable resource that all teachers have is each other. Without collaboration, our growth is limited to our own perspectives.”

Robert John Meehan

During the past four years, I have had the great fortune of being surrounded by many inspiring and supportive people. Each of them is a teacher to me, and without their help, this thesis would not be the same. I am very thankful for the opportunity to collaborate and learn from so many incredible people.

My biggest gratitude goes to my advisor, Prof. Dr. Stefan Ufer, who gave me this chance, supported me with his comprehensive knowledge and invaluable feedback, and always had an open door for me. Learning from his expertise and listening to his perspectives made me grow and was an experience that I will never forget. It was a pleasure to be his doctoral candidate and an honor to work in and with his team, for which I am very grateful.

I want to thank Prof. Dr. Andreas Obersteiner and Prof. Dr. Knut Neumann for joining the examination committee, and Prof. Dr. Konstantinos Panagiotou for chairing it.

I want to thank Prof. Dr. Daniel Sommerhoff, who supported me from the beginning as a critical friend and mentor. I am also very grateful for having had the opportunity to visit Prof. Dr. Jan Plass and his team at New York University.

I want to thank Deutsche Forschungsgemeinschaft (DFG) for funding this research (grant number UF59/5-1 and UF59/5-2, FOR 2385) and Prof. Dr. Thomas Augustin and the mentoring program of the faculty of Mathematics, Informatics, and Statistics at Ludwig-Maximilians-Universität München for supporting me during my PhD. I want to thank the whole COSIMA research group, my coauthors Prof. Dr. Kathleen Stürmer, Prof. Dr. Christof Wecker, Prof. Dr. Matthias Siebeck, and Maike Achtner, all my colleagues, all assistants, and all pre-service teachers participating in our simulations.

I want to thank my family and friends. You are my backbone. I want to thank my parents, Melanie and Rüdiger, and my siblings, Franziska, Katharina, and Philipp, for their unconditional love and support, and my niece Julia and nephews Benedikt and Luis for showing me how precious time is and how it flies. I want to thank Pia and Christian, who always believed in me and gave me strength with their boundless optimism and support. I am grateful for the empathic support of all my friends.

Citing William H. McRaven, just “Don’t ever, ever ring the bell.”

Summary

Teachers' diagnostic competence is essential for the effectiveness of their educational decisions. Facing the increasing heterogeneity of the student population, teachers' diagnostic competence becomes crucial to identifying the different needs of their students individually. Hence, teacher education programs aim to foster diagnostic competence. Accordingly, research interest in teachers' diagnostic competence has increased, aiming to derive implications for teacher education programs. However, a systematic investigation of the construct of diagnostic competence is scarce.

The research presented in this dissertation provides first steps toward such an investigation by approaching diagnostic competence using different indicators assessed close to real life in a developed simulation of diagnostic one-on-one interviews. The relationships between cognitive resources, affective-motivational resources, characteristics of the diagnostic process, and the quality of diagnostic performance are empirically investigated, measured in an authentic yet controlled setting. It was investigated whether cognitive and affective-motivational resources interact, how characteristics of the diagnostic process can be operationalized, and how the different indicators relate to each other. Additionally, it was investigated whether the presentation format of the simulation or the participants' perception of the simulation affect the development of pre-service teachers' interest.

This dissertation summarizes the findings of three manuscripts and reports on data of two studies. The first and second manuscripts refer to the first study ($N = 65$, $N = 63$) addressing the investigation of relationships between different indicators of diagnostic competence. The third manuscript reports on data from the second study ($N = 81$) analyzing the effects of the presentation format and perception of the simulation on the development of individuals' interest, comparing a role-play and a video simulation.

The results indicate that research and teacher training programs on diagnostic competence should address all indicators of diagnostic competence, namely cognitive and affective-motivational resources, characteristics of the diagnostic process, and the quality of diagnostic performance, since those indicators show significant relations and interactions. Addressing competence close to real life using simulations appears to be beneficial from an assessment and learning perspective.

Zusammenfassung

Die diagnostische Kompetenz von Lehrkräften ist wichtig für die Wirksamkeit ihrer pädagogischen Entscheidungen. Angesichts einer zunehmenden Heterogenität der Schülerpopulation ist die diagnostische Kompetenz von Lehrkräften von entscheidender Bedeutung, um die unterschiedlichen Bedürfnisse einzelner Schülerinnen und Schüler identifizieren zu können. Ein Ziel der Lehramtsausbildung ist daher die Förderung der diagnostischen Kompetenz. Dementsprechend hat auch das Forschungsinteresse an der diagnostischen Kompetenz von Lehrkräften zugenommen, mit dem Ziel, Empfehlungen für die Lehramtsausbildung abzuleiten. Eine systematische Untersuchung des Konstrukts der diagnostischen Kompetenz ist jedoch kaum vorhanden.

Die in dieser Dissertation vorgestellten Forschungsarbeiten stellen erste Schritte in Richtung einer solchen Untersuchung dar, indem diagnostische Kompetenz anhand verschiedener Indikatoren operationalisiert wird, die in einer entwickelten Simulation diagnostischer Einzelinterviews realitätsnah gemessen werden. Es wurden die Zusammenhänge zwischen kognitiven Ressourcen, affektiv-motivationalen Ressourcen, Merkmalen des diagnostischen Prozesses und der Qualität der diagnostischen Leistung, gemessen in einem authentischen, aber kontrollierten Setting, empirisch untersucht. Es wurde untersucht, ob kognitive und affektiv-motivationale Ressourcen interagieren, wie Merkmale des diagnostischen Prozesses operationalisiert werden können und wie sich die verschiedenen Indikatoren diagnostischer Kompetenz zueinander verhalten. Darüber hinaus wurde untersucht, ob das Präsentationsformat der Simulation oder die Wahrnehmung der Simulation die Entwicklung des Interesses der angehenden Lehrkräfte beeinflusst.

Die vorliegende Dissertation fasst die Ergebnisse von drei Manuskripten zusammen und berichtet Daten aus zwei Studien. Das erste und das zweite Manuskript beziehen sich auf die erste Studie $(N = 65, N = 63)$, die sich mit der Untersuchung von Beziehungen zwischen verschiedenen Indikatoren der diagnostischen Kompetenz befasst. Das dritte Manuskript bezieht sich auf Daten einer zweiten Studie $(N = 81)$, in der Effekte des Präsentationsformats der Simulation und der Wahrnehmung der

Simulation auf die Entwicklung des Interesses der Teilnehmenden untersucht wurden, wobei eine Rollenspiel- und eine Videosimulation verglichen wurden.

Die Ergebnisse deuten darauf hin, dass Forschungsarbeiten und Aus- beziehungsweise Weiterbildungsprogramme für Lehrkräfte zur diagnostischen Kompetenz alle Indikatoren der diagnostischen Kompetenz berücksichtigen sollten, genauer gesagt kognitive und affektiv-motivationale Ressourcen, Merkmale des diagnostischen Prozesses und die Qualität der diagnostischen Leistung, da diese Indikatoren signifikante Beziehungen und Interaktionen aufweisen. Die realitätsnahe Auseinandersetzung mit Kompetenz mittels Simulationen erscheint sowohl hinsichtlich der Messung, aber auch der Förderung von Kompetenzen vorteilhaft.

Included Manuscripts

The present dissertation is a cumulative thesis and comprises three published manuscripts. The first manuscript was published in an international peer-reviewed journal, and the second manuscript appeared in a national peer-reviewed journal in the English language. The third manuscript has been published in international peer-reviewed conference proceedings. The author of this dissertation was the first author of all manuscripts and carried the main responsibility for conceptualization, data analyses, writing the original and revised manuscripts and pursuing the publication process.

Manuscript A

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, 6:604568. https://doi.org/10.3389/feduc.2021.604568

The first author (Stephanie Kron) was primarily responsible for the publication of this manuscript. The coauthors, Prof. Dr. Daniel Sommerhoff, Maike Achtner, and Prof. Dr. Stefan Ufer, contributed to the development and publication of the manuscript with critical reviews.

Manuscript B

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*, 135-172. https://doi.org/10.1007/s13138-022-00200-2

The first author (Stephanie Kron) was primarily responsible for the publication of this manuscript. The coauthors, Prof. Dr. Daniel Sommerhoff, Maike Achtner, Prof. Dr. Kathleen Stürmer, Prof. Dr. Christof Wecker, Prof. Dr. Matthias Siebeck, and Prof. Dr. Stefan Ufer, contributed to the development and publication of the manuscript with critical reviews.

Manuscript C

Kron, S., Sommerhoff, D., Achtner, M., Stürmer, K., Wecker, C., Siebeck, M., & Ufer, S. (2022). Simulation-based learning environments: Do they affect learners` relevant interests? In C. Fernández, S. Llinares, Á. Gutiérrez, & N. Planas (Eds.), Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education (Vol. 3, pp. 67-74). PME. https://rua.ua.es/dspace/bitstream/10045/126600/1/proceedings-pme-45-vol3-09.pdf

The first author (Stephanie Kron) was primarily responsible for the publication of this manuscript. The coauthors, Prof. Dr. Daniel Sommerhoff, Maike Achtner, Prof. Dr. Kathleen Stürmer, Prof. Dr. Christof Wecker, Prof. Dr. Matthias Siebeck, and Prof. Dr. Stefan Ufer, contributed to the development and publication of the manuscript with critical reviews.

Content

1 Introduction

Teachers' diagnostic competence is seen as a fundamental factor in successful teaching (Herppich et al., 2018; Leuders et al., 2018). Students' individual needs can only be addressed adaptively if their teachers can accurately and efficiently diagnose those needs. Considering the relevance of diagnostic competence for teaching and teacher training, several research projects have investigated diagnostic competence, aiming to develop models of diagnostic competence and derive implications for fostering (pre-service) teachers' diagnostic competence. This dissertation contributes to this endeavor by investigating the construct of diagnostic competence through the relationships between different indicators of diagnostic competence. Following the call to assess competence close to real-life settings (e.g., Shavelson, 2012; Albu & Lindmeier, 2023), a simulation of diagnostic one-on-one interviews was developed to measure indicators of pre-service teachers' diagnostic competence. Understanding a diagnosis as a data-based result of a systematic, reflective process (Helmke, 2009), indicators of the diagnostic process and diagnostic performance are assessed.

Referring to general theories of professional competence, competence builds on (to be) acquired cognitive resources that are needed to solve a specific problem, as well as motivational, volitional, and social capabilities that are required for successful problem solving in variable situations (Weinert, 2001). Following Blömeke et al. (2015), competence is a continuum between an individual's resources and their performance in professional situations. Current conceptualizations of professional competence propose capturing competence not only by individuals' resources and performance but also by their processes during professional demands preceding their performance (Hammer & Ufer, 2023).

For diagnostic competence, various conceptualizations are put forth by different research projects. The research group *NeDiKo* (Herppich et al., 2018) understands diagnostic competence as a cognitive disposition that is reflected in diagnostic activities during diagnostic situations and filtered by an individual's motivational resources and processes, resulting in a diagnosis that itself results in educational decisions. The research group *COSIMA* (Heitzmann et al., 2019) defines diagnostic competence as "[...] individual dispositions enabling people to apply their knowledge in

diagnostic activities according to professional standards to collect and interpret data in order to make high-quality decisions" (Heitzmann et al., 2019, p. 5).

The *COSIMA* research group investigates diagnostic competence using simulation-based learning environments. This approach aims to contribute to the understanding of diagnostic competence by assessing this construct in highly authentic, yet controlled, situations, and to gain insights regarding the fostering of diagnostic competence using simulation-based learning environments (Fischer & Opitz, 2022).

The *COSIMA* framework (see Figure 1) illustrates this research approach. Diagnostic competence is assumed to be reflected in an individual's professional knowledge base, their diagnostic activities (e.g., questioning, generating hypotheses, evaluating evidence), and the quality of their diagnosis (e.g., accuracy and efficiency). In terms of fostering diagnostic competence, an individual's cognitive and affective-motivational learning prerequisites need to be considered. These learning prerequisites are assumed to affect an individual's learning in simulations. Regarding the characteristics of the simulation-based learning environment itself, the effects of instructional support (e.g., explicit presentation of information, scaffolding) are of special interest. The effects of contextual factors, such as characteristics of the diagnostic situation (e.g., individual versus collaborative setting, interaction versus document-based diagnosis) and the domain (i.e., teacher versus medical education), are also investigated.

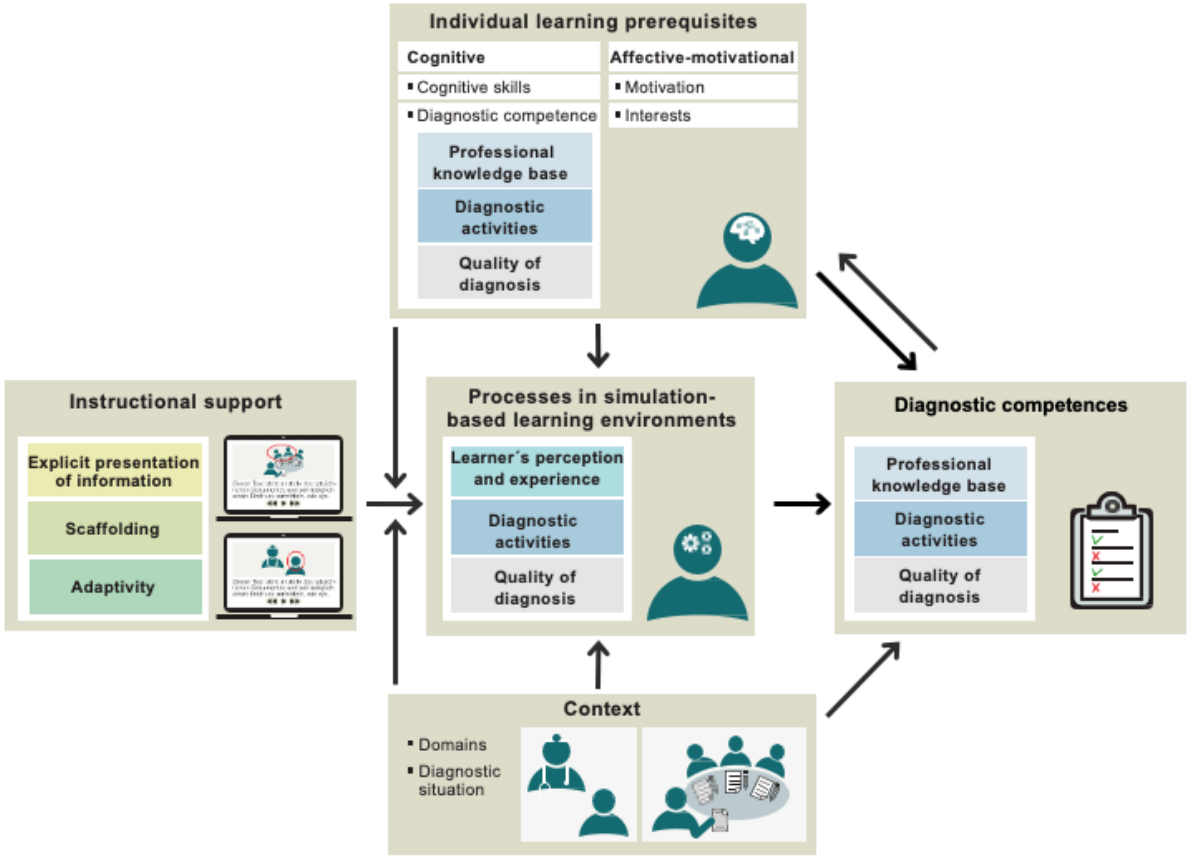

Figure 1: Conceptual framework of the research group COSIMA[1]

Connecting the theoretical implications of competence as a continuum (Blömeke et al., 2015), the *NeDiKo* (Herppich et al., 2018), and the *COSIMA* (Heitzmann et al., 2019) framework, this dissertation conceptualizes diagnostic competence using different indicators, namely individuals' resources, characteristics of the diagnostic process, and the quality of diagnostic performance. Three research aims are addressed (see Figure 2): following the definition of Heitzmann et al. (2019), the relationships of pre-service teachers' cognitive resources (i.e., professional knowledge) to process and performance indicators of diagnostic competence are investigated (relationships between cognitive resources, diagnostic processes, and diagnostic performance). Taking the perspective of affective-motivational resources moderating the effects of individuals' cognitive resources (Herppich et al., 2018; Blömeke et al., 2015), the effects of pre-service teachers' affective-motivational resources (i.e., interest) interacting with professional knowledge on the diagnostic process and diagnostic performance were investigated (the effects of the interaction of cognitive and affective-

[1] COSIMA framework model (https://www.for2385.uni-muenchen.de/aktuelles/rahmenmodellccby/cosima-frame-model_eng.pdf), COSIMA research unit, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/legalcode)

motivational resources on the diagnostic process and diagnostic performance). Moreover, it was investigated whether learning experiences made in different simulation-based learning environments affect the development of individuals' resources (i.e., interest).

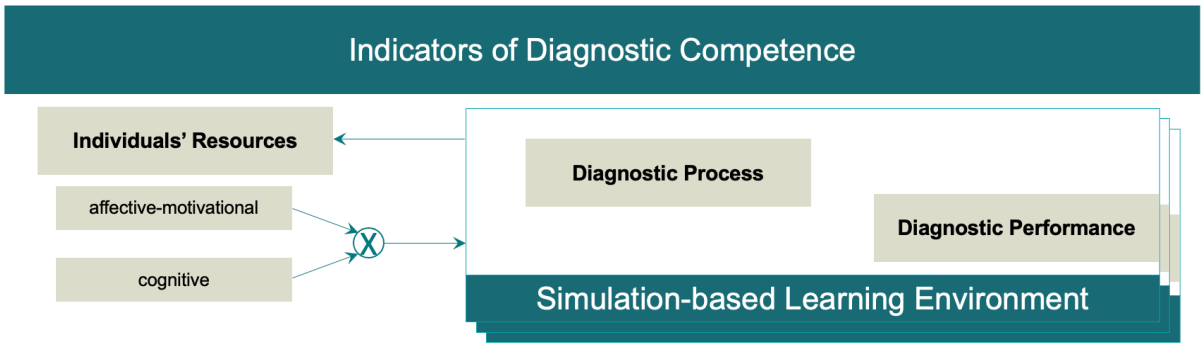

Figure 2: Adapted excerpt from the COSIMA framework underlying this dissertation

2 Theoretical Background

2.1 Indicators of Diagnostic Competence

Going beyond the perspective of Blömeke et al. (2015), this dissertation approaches diagnostic competence not only by assessing the individuals' resources and their diagnostic performance but also by investigating indicators of the diagnostic process itself (cf. Hammer & Ufer, 2023). The set of indicators presented strives to be used as a starting point for the assessment and fostering of diagnostic competence.

2.1.1 Individuals' Resources

Several definitions of diagnostic competence conceptualize competence by an individual's cognitive resources (e.g., Klieme & Leutner, 2006; Weinert, 2001; Blömeke et al., 2015; Heitzmann et al., 2019; Herppich et al., 2018). In the context of teacher education, cognitive resources are commonly investigated, focusing on professional knowledge. The professional knowledge of pre-service teachers is mostly conceptualized following Shulman's (1987) categorization, distinguishing content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical knowledge (PK). With respect to mathematics education, teachers' CK commonly refers to knowledge about school mathematics and their mathematical background (Baumert & Kunter, 2013). However, the term CK is sometimes defined as knowledge about academic mathematics, whereas knowledge about school-related mathematics is termed school-related content knowledge (SRCK) (e.g., Kleickmann et al., 2014). For the research presented in this dissertation, CK is understood as defined by Baumert and Kunter (2013). Mathematics teachers' PCK includes subject-specific knowledge about the teaching and learning of mathematics, such as knowledge about student cognition, diagnostic tasks, or instructional strategies (Baumert & Kunter, 2013; Depaepe et al., 2013). In contrast, teachers' PK commonly refers to knowledge about the general aspects of teaching and learning (Kleickmann et al., 2014). Several studies in the context of diagnostic competence provide evidence that the different components of professional knowledge (i.e., CK, PCK, PK) relate to different aspects of diagnostic competence, such as diagnostic activities or diagnostic accuracy (e.g., Binder et al., 2018; Kramer et al., 2021; Nickl et al., 2022; Ostermann, 2018; van den Kieboom et al., 2014). Despite increasing research interest in the field of diagnostic

competence, systematic investigations of the specific relations between single components of professional knowledge and indicators of the diagnostic process and performance are lacking.

Before accessible professional knowledge is applied in a professional situation, this knowledge needs to be activated. Herppich et al. (2018) propose that the activation of relevant cognitive dispositions is moderated by other dispositions, such as individuals' trait interest as an affective-motivational resource. For knowledge to be activated in a specific situation, it is not sufficient that the knowledge has already been acquired; the particular knowledge needs to be identified as relevant to the specific situation. In this context, Hidi (1990) suggests that an individual's interest is crucial for the activation of knowledge, for example, by contributing to sustained attention (Ainley et al., 2002). Individual interest as a stable person–object relation is termed the "tendency to occupy oneself with an object of interest" (intrinsic component; Krapp, 2002), which may also increase the tendency to activate respective knowledge, like knowledge about diagnosis or mathematics education, in specific situations. The perspective of interest as a person–object relation implies that levels of interest differ in accordance with different objects (e.g., a certain knowledge domain or a certain professional situation, such as educational diagnosis). This suggests that the activation of certain components of professional knowledge may relate to different facets of interest (e.g., interest in diagnosis, interest in mathematics education).

2.1.2 Quality of Diagnostic Performance

Referring to competence as a continuum as defined by Blömeke et al. (2015), competence refers to an individual's coherent performance over a range of specific situations in a certain period of time of similar quality and in relation to specified criteria. Heitzmann et al. (2019) argued that diagnostic competence, mirrored in a diagnostic judgment, is the basis for educational decisions. Claiming for appropriate educational decisions, the quality of the diagnostic judgment becomes crucial. The most prominent measure used to draw conclusions about an individual's diagnostic competence is the accuracy of the diagnostic judgement (Chernikova et al., 2020a; Chernikova et al., 2020b). Accuracy describes the match between teachers' judgments of students' understanding and their actual understanding, as assessed by an independent test (Südkamp et al., 2012). If we understand diagnostic judgments as the basis for educational decisions, then more accurate diagnoses can be assumed to lead to

accurate educational decisions, which may relate to higher qualities in student learning. Indeed, Behrmann and Souvignier (2013) report evidence for the relationship between teachers' diagnostic accuracy and their students' achievement.

However, solely focusing on the accuracy of a diagnostic judgment as an indicator of diagnostic competence has been criticized for not taking the underlying diagnostic process into account (Herppich et al., 2018). Addressing this criticism, this dissertation also investigates the characteristics of the diagnostic process as an indicator of diagnostic competence.

2.1.3 Characteristics of the Diagnostic Process

To investigate diagnostic processes, different models have been established. Generic models focusing on process regulation (Klug et al., 2013), as well as finer-grained models investigating how diagnostic evidence is gathered and processed (Heitzmann et al., 2019; Herppich et al., 2018; Loibl et al., 2020). In the *NeDiKo* model (Herppich et al., 2018), the diagnostic process subsumes prototypical decisions and diagnostic actions aiming to gather information about students' understanding. An important element here is teachers' ability to notice and describe relevant events in classroom situations, also known as professional vision (Kramer et al., 2020; Seidel & Stürmer, 2014). Heitzmann et al. (2019) define diagnosing as the goal-directed accumulation and integration of information to reduce uncertainty regarding educational decisions. These diagnostic activities aim to gather and evaluate diagnostic evidence, generate and test alternative hypotheses, and draw conclusions for practice (e.g., educational or medical decisions) from this evidence (Heitzmann et al., 2019). This means that every diagnostic process relies on diagnostic activities, such as questioning, hypothesis generation, and the generation and evaluation of evidence.

Regarding mathematics education, such diagnostic evidence about students' understanding is commonly drawn from students' responses to tasks (Black & Wiliam, 2009; Schack et al., 2013). Focusing on the use of tasks to stimulate student responses to be used as diagnostic evidence, the selection of suitable tasks triggering student responses that reveal relevant diagnostic information becomes crucial. In this dissertation, a task's potential to stimulate diagnostically relevant and reliable information about a student's mathematical understanding is termed "diagnostic potential." The selection of tasks with high diagnostic potential during a diagnostic

process is thus seen as a process indicator of diagnostic competence. Additionally, making adaptive use of a task's diagnostic potential is assumed to be a second process indicator of diagnostic competence, in the sense that pre-service teachers adjust their selection of tasks not only by the task's diagnostic potential but also by the diagnostic evidence that could have been gathered from prior tasks.

The investigation of relations between individuals' resources, indicators of the diagnostic process, and indicators of diagnostic performance (see Figure 2), aims to contribute to a deeper understanding of diagnostic competence. The operationalization of pre-service teachers' task selection as a process indicator of diagnostic competence broadens the research field of diagnostic competence by not solely focusing on the diagnostic judgment itself. This approach can give access to the extent to which individuals' resources affect performance directly or indirectly via process measures. However, this requires validation, addressing the question of which process measures—beyond individuals' resources—are predictive for diagnostic performance (e.g., accuracy). Indeed, this validation goes beyond the scope of this dissertation. By investigating the interaction effects between certain individuals' resources on the diagnostic process and performance, this dissertation aims to provide further empirical evidence for the role of cognitive resources (e.g., Kramer et al., 2021; Nickl et al., 2022) and insights about the role of affective-motivational resources. More precisely, this dissertation addresses the research gap evolved by the assumption proposed by Herppich et al. (2018) that the effects of cognitive resources are moderated by affective-motivational resources. Finally, the investigation of the effects of the presentation format of the simulation and its perception on the development of individuals' affective-motivational resources provides insights regarding the design of simulations used to assess or foster professional competence.

Conceptualizing diagnostic competence by different indicators considering resources, processes, and performance allows for a holistic assessment of diagnostic competence. This dissertation strives to contribute to the understanding of the interplay between individuals' resources, diagnostic processes, and diagnostic performance as a starting point for the development of a simulation-based learning environment used to foster diagnostic competence. A deeper understanding of the relationships between the different indicators of diagnostic competence may provide insights into potential

levers for in-process scaffolding to support diagnostic activities and the development of diagnostic competence.

2.2 Using Simulation-Based Learning Environments to Assess and Foster Diagnostic Competence

Understanding competence as a "[...] complex ability [...] closely related to performance in real-life situations" (Hartig et al., 2008, p. v), research emphasizes an assessment of professional competence close to real-life situations (Kaiser & König, 2019). Following Shavelson (2010), the observation of competence demands tasks that should "[…] (a) be real-life in nature, (b) tap complex abilities and skills, (c) be amenable to practice and improvement, and (d) be amenable to standard setting." Accordingly, besides the criteria of reliability and validity, multiple quality criteria for assessment formats of competence, such as authenticity, fairness, transparency, and cost efficiency, have been discussed (Blömeke et al., 2015; Messick, 1995). Ufer and Neumann (2018) argue that the assessment of competence requires assessment formats which go beyond the measurement of declarative knowledge by demanding individuals to "[…] link multiple abilities and skills […]" (Ufer & Neumann, 2018, p. 437).

To address these requirements, the so-called approximations of practice (AoPs; Grossman et al., 2009) are discussed. By reconstructing real-life situations, for example, using simulations, AoPs allow for the application of professional knowledge and the assessment of competence close to real-life situations (e.g., Shavelson, 2010; Albu & Lindmeier, 2023). Since AoPs only approximate real-life situations, the design of an AoP allows us to focus on certain aspects of professional demands, to control disruptive factors, and to limit chosen features that would increase the complexity of the respective situation. This possibility of controlling the characteristics of the simulated real-life situation contributes to the fairness and transparency of the assessment format, as identical situations can be presented to several individuals. Moreover, AoPs are also replicable, allowing for repeated measurements to investigate whether an individual's competence develops over time.

The possibility of controlling certain circumstances of the simulated real-life situation allows the application of professional knowledge in situations that are less overwhelming compared to the real-life situation (Codreanu et al., 2020). Thus, using AoPs as a learning environment is also discussed, as they may cause a lower cognitive

load compared to the respective real-life situation (Sweller, 2010). Moreover, AoPs allow learners to apply their acquired knowledge in an authentic, yet controlled, setting and offer the possibility of implementing specific scaffolds aimed at supporting the development of the respective competence, for example, by providing instructional support or conceptual prompts (Irmer et al., 2022; Sommerhoff et al., 2023). Therefore, especially in the context of higher education (e.g., Stegmann et al., 2012), AoPs are discussed as promising learning environments.

Regardless of the purpose for which an AoP is used (e.g., assessment or fostering), Grossman et al. (2009) argue that, for the design of an AoP, it is central to build real-life demands that are highly authentic, allowing the individual to immerse themselves in the simulated situation (Goeze et al., 2014). Chernikova et al. (2023) subdivide authenticity into physical resemblance and functional correspondence. Physical resemblance refers to surface elements of the simulation that are easily observable, whereas functional correspondence describes the match between the simulated situation and the real-life job demand on a task level (Chernikova et al., 2023; Stürmer et al., under review). Although functional correspondence is discussed as being more important than physical resemblance when using simulations to foster competence development (e.g., Hamstra et al., 2014), empirical evidence is scarce. Admittedly, designing highly authentic simulations can conflict with the aim of reducing the complexity of a professional situation with regard to cognitive load theory (Sweller, 2010). Citing Codreanu et al. (2020), the challenge here is to find a balance between the level of authenticity of the simulated situation and the cognitive demand increasing when designing highly authentic situations.

To simulate professional job demands as an AoP, different presentation formats of simulations can be used, differing in their balance between authenticity (i.e., physical resemblance and functional correspondence) and complexity. Common presentation formats are role-play simulations, which are often used in medical education (e.g., Stegmann et al., 2012), and video-based formats, which are popular in teacher education (e.g., Seidel et al., 2011). Role-play simulations allow the development of highly authentic situations in the sense of functional correspondence, as they offer more opportunities for interaction so that almost any professional task can be simulated (e.g., freely asking verbal probing questions in a live role-play simulation of one-on-one interviews). Hence, they are assumed to be perceived as more authentic

and immersive but are in danger of provoking higher cognitive load due to the complexity caused by the variety of options for interaction. Conversely, the development of video-based simulations inherently requires the definition of concrete structures, offering fewer opportunities for interaction or self-regulated actions of the learners, as they are based on pre-defined options for action (e.g., pre-defined probing questions in a video simulation of one-on-one interviews), which may lead to a lower perception of authenticity due to the lower functional correspondence, but also to a lower extraneous cognitive load (e.g., pre-defined and limited interaction). Depending on the real-life situation being approximated, video simulations can be beneficial in terms of physical resemblance. In AoPs in which students (e.g., 6th graders) have to be simulated, role-plays are reliant on adult actors (e.g., ethical reasons, extensive training necessary to guarantee repeatability and validity of the simulated situation), whereas videos of real students (i.e., picture and voice of a real 6th grader) can be used for video simulations.

Research on whether, and to what extent, the presentation format of a simulation affects the individual's perception of the simulation and the development of indicators of competence is scarce. To address this gap, this dissertation investigates the effects of a simulation's presentation format and the perception of the simulation on the development of pre-service teachers' relevant interests.

3 Research Questions and Overview of the Included Manuscripts

The aim of this dissertation is to contribute to the understanding of pre-service secondary school mathematics teachers' diagnostic competence, as well as to the role of simulations when assessing and fostering pre-service teachers' diagnostic competence during their teacher education program. This endeavor is split into three main goals: (1) operationalizing task selection as a process indicator of diagnostic competence by investigating to what extent the concept of selecting tasks having high diagnostic potential can be observed in pre-service teachers' task selection during simulated diagnostic interviews; (2) investigating the role of individuals' resources (interest and professional knowledge, particularly in their interaction) for the enactment of diagnostic processes and for diagnostic performance; as well as switching from the assessment perspective to the learning perspective, (3) investigating the role of the simulation's presentation format and the perception of the diagnostic simulation for the development of individuals' interest. Each of these goals is addressed by at least one of the three manuscripts included in this dissertation.

Manuscript A addresses the first (1) and second (2) goals. In this manuscript, the construct of a task's diagnostic potential being part of teachers' PCK (Baumert & Kunter, 2013) is investigated, observing pre-service teachers' task selection as a process indicator of diagnostic competence. Addressing the second goal, it was investigated how single components of pre-service teachers' professional knowledge relate to the selection of tasks with high diagnostic potential (Baumert & Kunter, 2013). Additionally, whether the selection of high-potential tasks was adapted to different understandings of the respective student that could have been extracted from prior tasks was investigated. The research questions addressed in Manuscript A are as follows:

RQ A.1: To what extent are pre-service teachers sensitive to the diagnostic potential of tasks? Is there systematic variation in pre-service teachers' sensitivity to diagnostic potential?

RQ A.2: To which extent is pre-service teachers' sensitivity to the diagnostic potential of tasks related to different components of their professional knowledge (CK, PCK, PK)?

RQ A.3: To which extent is pre-service teachers' task selection adaptive to evidence generated from prior tasks? Is there systematic variation in pre-service teachers' adaptive use of diagnostic task potential?

RQ A.4: To which extent is pre-service teachers' adaptivity in the use of diagnostic task potential related to different components of their professional knowledge (CK, PCK, PK)?

Manuscript B builds on the results of Manuscript A, that it was pre-service teachers' CK, but not their PCK, which was related to the selection of tasks having high diagnostic potential (for a detailed summary, see Section 5.1) and the theoretical assumptions about the interplay of cognitive and affective-motivational resources (Herppich et al., 2018), investigating the main and interaction effects of pre-service teachers' professional knowledge components and their interest. In accordance with the person–object theory of interest (Krapp, 2002), two facets of interest were distinguished (interest in diagnosis, interest in mathematics education). Besides the process indicator of task selection, diagnostic accuracy as a performance indicator was also considered, contributing to the first (1) and second (2) goals of this dissertation. The research questions addressed in Manuscript B are as follows:

RQ B.1: How are the different components of pre-service teachers' professional knowledge (CK, PCK, PK) related to pre-service teachers' sensitivity to diagnostic task potential and diagnostic accuracy?

RQ B.2: How are different facets of pre-service teachers' interest (interest in diagnosis, interest in mathematics education) related to pre-service teachers' sensitivity to diagnostic task potential and diagnostic accuracy?

RQ B.3a: Does the interaction of professional knowledge and interest affect pre-service teachers' sensitivity to diagnostic task potential and their diagnostic accuracy?

RQ B.3b: Which knowledge components and interest facets interact in their effects on pre-service teachers' sensitivity to diagnostic task potential and diagnostic accuracy?

Manuscript C addresses the third (3) goal, building on the results of Manuscript B, which confirm the assumption of affective-motivational resources filtering the effects of cognitive resources, observing significant interaction effects between pre-service teachers' interest and their professional knowledge on process and performance indicators of diagnostic competence (for a detailed summary, see Section 5.2) and implications from theory regarding the development of interest (e.g., Rach, 2021). It was investigated whether the simulation's presentation format or the participants' perception of the simulation affect the development of participants' interest. The research questions addressed in Manuscript C are as follows:

RQ C.1: Does the presentation format of a simulation-based learning environment affect participants' relevant situational interests reported after the simulation?

RQ C.2: How do participants' initial individual interests and their perception of the simulation affect participants' situational interests after the simulation?

RQ C.3: Does the presentation format influence the development of situational interest over multiple simulations, after controlling for the perception of the simulation?

For a summary of the results of the research questions of Manuscript C, see Section 5.3.

The research questions addressed in Manuscripts A and B were investigated using data from the same study, in which the assessment of indicators of diagnostic competence was focused on to contribute to the understanding of the specific roles of single professional knowledge components for individuals' diagnostic processes and diagnostic performance. The research questions in Manuscript C refer to a second study in which the fostering of diagnostic competence was pursued (see Figure 3), aiming to contribute to the understanding of the potential of simulations to affect individuals' affective-motivational resources.

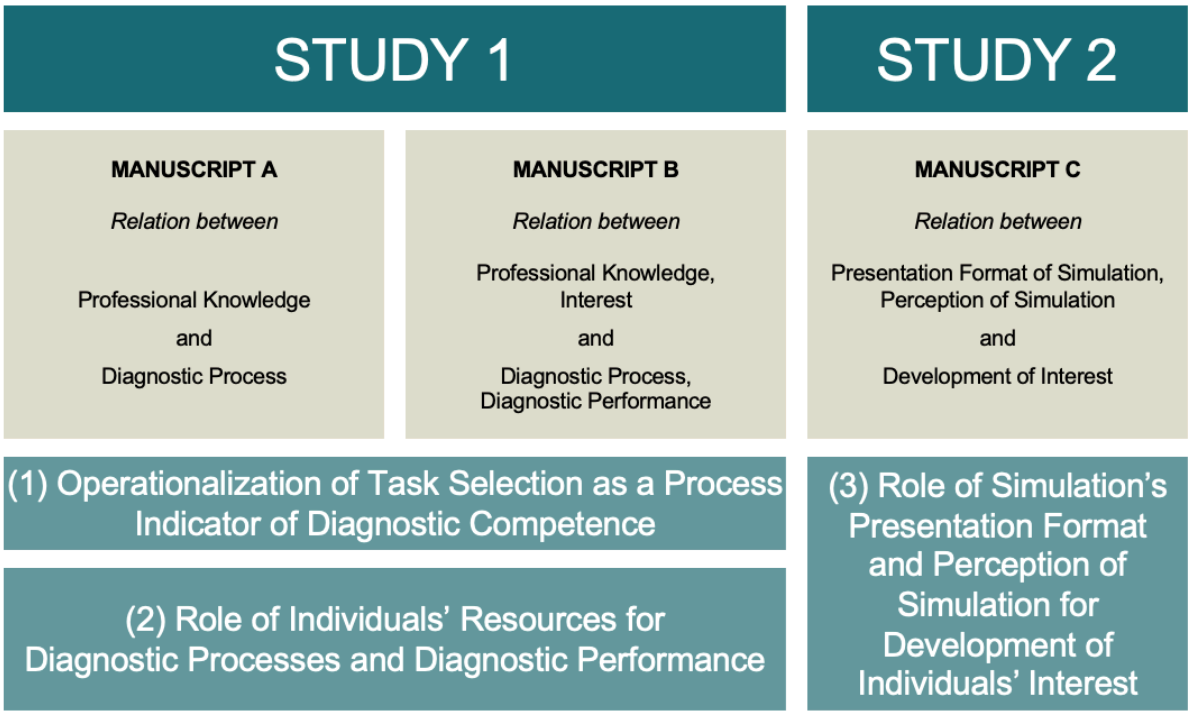

Figure 3: Overview of included manuscripts and addressed goals of this dissertation

4 Methodology

4.1 The Research Project DiMaL as Part of the COSIMA Research Group

The research presented in this dissertation is part of the research project DiMaL[2] (UF59/5-1 and UF59/5-2), which is part of the DFG research unit 2385 COSIMA[3], funded by the Deutsche Forschungsgemeinschaft (DFG). Consisting of eight research projects, COSIMA investigates the assessment and fostering of diagnostic competence using simulations in the context of teacher education and medical education.

As one of five research projects located in the field of teacher education, the DiMaL research project aims to contribute to the understanding of pre-service secondary school mathematics teachers' diagnostic competence. In detail, the simulation-based assessment of different indicators of diagnostic competence, covering individuals' resources, the diagnostic process, and diagnostic performance, is investigated. The aim of this research project is to gain insights into the specific relations between these indicators of diagnostic competence (e.g., being a predictor versus a moderator variable). Moreover, the effects of the presentation format of the simulation and the individuals' perception of the simulation on the development of the individuals' resources (i.e., interest) are investigated. Building on the insights of this dissertation, several intervention studies in the DiMaL project currently investigate different scaffolding measures (e.g., reflection support, knowledge activation, role-taking) to foster diagnostic competence, as well as their differential effectiveness depending on individuals' resources.

To investigate the assessment and development of indicators of diagnostic competence, a simulation of one-on-one diagnostic interviews was developed. To allow for an investigation of the effects of the simulation's presentation format, the simulation was implemented as a live role-play simulation and a video-based simulation. Both presentation formats of the simulation were implemented in the same

[2] The acronym *DiMaL* refers to the diagnostic competence of mathematics teachers (Diagnose mathematischer Lernvoraussetzungen).
[3] For more information, see www.for2385.uni-muenchen.de

web-based system. The system guides the participants through the whole simulation without any need for external support.

4.2 Design of the Simulations

During the simulation, the participants (pre-service teachers) take over the role of a teacher, aiming to diagnose a simulated 6th grader's understanding of decimal fractions[4] in a one-on-one interview. The participants are asked to select diagnostic tasks from a given task set, observe the simulated student's solution, make notes during the diagnostic interview, and then diagnose the student's mathematical understanding of decimal fractions. The task set that is provided for the diagnostic one-on-one interviews consists of 45 mathematical tasks[5] relating to decimal fractions. These tasks vary in their diagnostic potential. The first time participants conduct an interview, they become familiar with the system itself and are instructed to analyze the task set with respect to the diagnostic potential of the single tasks. The interview phase follows directly after the familiarization phase and is limited to a maximum duration of 30 minutes[6]. After the interview, the participants are asked to diagnose the interviewed student's mathematical understanding in an open report and in a closed rating (see Figure 4). Eight experts from the field of teacher education or research on teacher education, and $N = 13$ pre-service teachers validated the simulation as an authentic simulation of real-life professional demands (Stürmer et al., 2021).

The simulation was established in two presentation formats, initially as a live role-play simulation in which trained research assistants play the role of the simulated 6th grader by following the instructions provided in the student case profiles (Marczynski et al., 2022). During this live role-play simulation, the participants in the teacher's role can ask any verbal question regarding the student's solution by directly interacting with the simulated student.

After establishing this role-play simulation and gathering some experiences about typical probing questions, a video-based version of the simulation was developed (Kron et al., submitted). For the video-based simulation, scripted videos of 6th graders

[4] According to the German curricula, decimal fractions are taught in grade 6.
[5] Only for the first study: Due to adaptations in design and structure, the task set for the second study was reduced to a total number of 35 tasks.
[6] Only for the first study: Due to adaptations in design and structure, the interview duration for the second study was shortened to 25 minutes.

for each single diagnostic task had been recorded, based on the student case profiles' descriptions used for the role-play simulation. To allow for probing questions in the video simulation, six to eight possible probing questions referring to the respective student case profile's initial task solution were defined, and videos containing the students' answers to all of those single probing questions were recorded.[7]

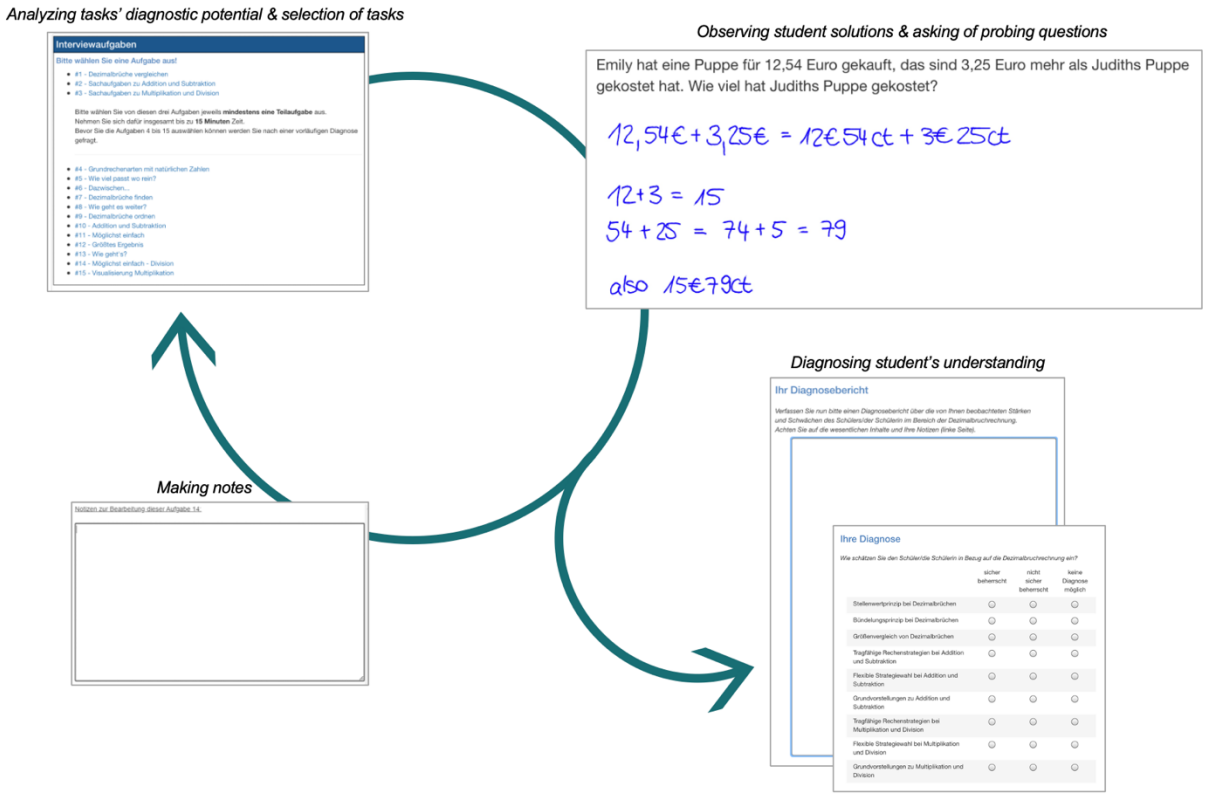

Figure 4: Illustration of the simulated professional demands

4.3 Mathematical Content of the Simulated Diagnostic One-on-One Interviews: Decimal Fractions

The topic of decimal fractions was selected as mathematical content for the simulations due to its importance in everyday life and the fact that several studies on students' understanding of decimal fractions report a variety of systematic student errors (e.g., Heckmann, 2006; Lortie-Forgues et al., 2015; Padberg & Wartha, 2017; Resnick et al., 1989; Steinle, 2004). Moreover, the teaching of decimals is a topic in university teacher training in several countries (e.g., Lortie-Forgues et al., 2015; Padberg & Wartha, 2017).

[7] See Appendix for screenshots of both presentation formats of the simulation.

In many cases, student errors in the area of decimal fractions occur systematically due to specific misconceptions regarding the place value system, the structure of decimal fractions, the meaning of basic arithmetic operations, or calculation strategies for the four basic arithmetic operations (Heckmann, 2006; Marczynski et al., 2022). Resnick et al. (1989) classified students' errors regarding the place value system and structuring of decimal fractions into three categories: whole number rule (WNR), fraction rule (FR), and zero rule (ZR). Students using the WNR interpret the decimal portion of the decimal fraction as a separate whole number (also known as "natural number bias", Alibali & Sidney, 2015; cf. Steinle & Pierce, 2006), FR refers to interpreting the decimal portion like fractions (i.e., knowing that thousandths are smaller parts than hundredths, that three-digit decimal fractions are read as thousandths, and that those are smaller than two-digit decimal fractions which are read as hundredths: also known as the "longer is smaller" strategy, Padberg & Wartha, 2017), and ZR labels the strategy of assuming decimal fractions with a zero following the decimal point to be smaller (Resnick et al., 1989; Steinle, 2004). The fact that these superficial strategies can lead to correct solutions in some mathematical tasks illustrates the importance of teachers' competence in identifying and diagnosing misconceptions by selecting appropriate mathematical tasks for their diagnosis. For example, a comparison of 0.417 being larger than 0.3 could even be solved correctly by students solely using the WNR (comparing 417 to 3). The comparison of 0.354 to 0.55 has a higher diagnostic potential, as it would lead to incorrect answers of students using the WNR (comparing 354 to 55), and therewith directly indicating a potential misconception. Misconceptions regarding the place value system can also cause errors when performing basic arithmetic operations (Heckmann, 2006). Regarding multiplication and division, calculations with rational numbers lead to new phenomena (e.g., multiplication can have a result that is smaller than one of the two factors, and division can have a result that is bigger than the dividend) (Fischbein et al., 1985; Siegler & Lortie-Forgues, 2015). Therefore, it is highly important that teachers are able to diagnose such misconceptions accurately at an early stage in order to make appropriate educational decisions.

Based on the variety of student misconceptions, it was possible to design four different student case profiles varying in their understanding in nine subareas of decimal fractions, namely (1) the place value principle, (2) the bundling principle, (3) the comparison of decimals, (4) the basic concepts of addition and subtraction, (5)

arithmetic abilities in addition and subtraction, (6) flexible use of strategies regarding addition and subtraction, (7) the basic concepts of multiplication and division, (8) arithmetic abilities in multiplication and division, and (9) flexible use of strategies regarding multiplication and division. This set of student case profiles allows for the arrangement of multiple sessions of the simulation to generate data for different measurement points.

4.4 Design of Studies and Procedure

The manuscripts included in this dissertation refer to two different studies. The research findings published in the first and second manuscripts relate to *Study 1*. This study aimed to investigate different indicators of diagnostic competence and the relationships between these indicators. After an initial familiarization phase, the participants of this study conducted two role-play simulations consecutively. Directly after the simulations, a paper-and-pencil test to assess professional knowledge followed, and demographic information was recorded (see Figure 5). The study was conducted in the winter term 2019/2020 at the Ludwig-Maximilians-Universität München. The data collection took place in person and had an average total duration of three hours per participant. The two student case profiles used for the two role-play simulations were assigned randomly and enacted by different trained research assistants. The sample of this study consists of $N = 65$ pre-service secondary school mathematics teachers[8] ($N = 130$ simulations in total).

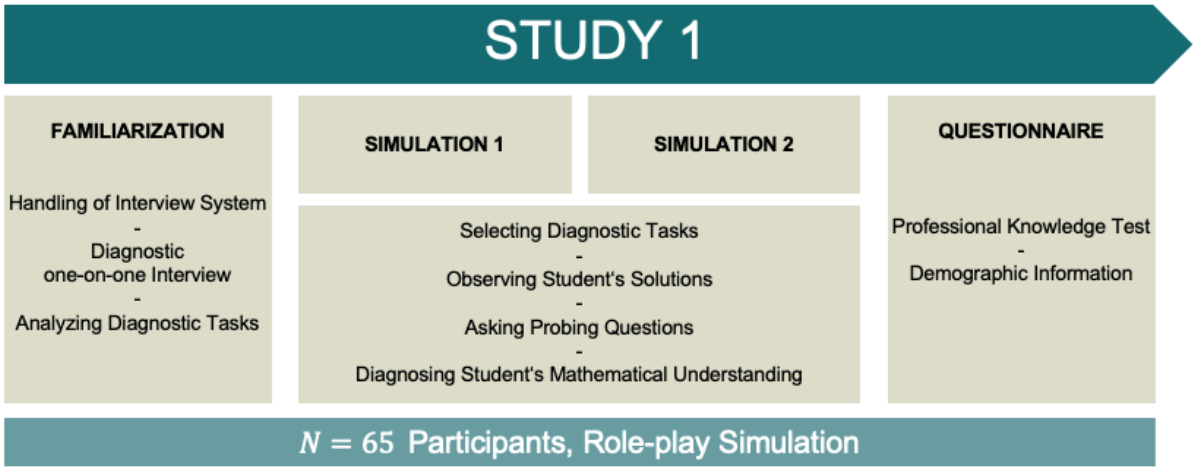

Figure 5: Design of the first study; findings reported in Manuscripts A and B

[8] For the data analysis of Manuscript B, two participants had been excluded, leading to a sample size of $N = 63$ participants.

The third manuscript included in this dissertation reports on the data of a second study, *Study 2*. In this study, the effects of the presentation format of a simulation-based learning environment and the perception of the simulation on the development of individuals' interest were investigated by sequencing four simulations and randomly assigning participants to the same presentation format (role-play or video simulation) for all four simulations. Between each of the two simulations was a timespan of two weeks. The familiarization phase was presented directly before conducting the first simulation. Data on the professional knowledge test (paper-and-pencil) and demographic information were collected two weeks before the first simulation (see Figure 6). Before the first simulation and after each simulation, the individuals' interest was assessed. The perception of the simulation was measured during the simulation. The study took place during the summer term of 2021 at the Ludwig-Maximilians-Universität München and was done online due to the COVID-19 pandemic. A single simulation had an average duration of one hour. The data collection for both presentation formats took place online using the online video conference software ZOOM[9]. For the role-play simulations, each participant and the research assistant enacting the student case profile met in a web conference room. To allow the participant to monitor the simulated student's solution of tasks, those solutions were handwritten on an external screen, which was shared by screen sharing. Each participant diagnosed four different student case profiles, which were randomly assigned to single simulations. $N = 39$ pre-service secondary school mathematics teachers conducted the role-play simulation; $N = 42$ conducted the video simulation ($N = 81$ participants and $N = 324$ simulations in total).

[9] https://explore.zoom.us/de/products/meetings/

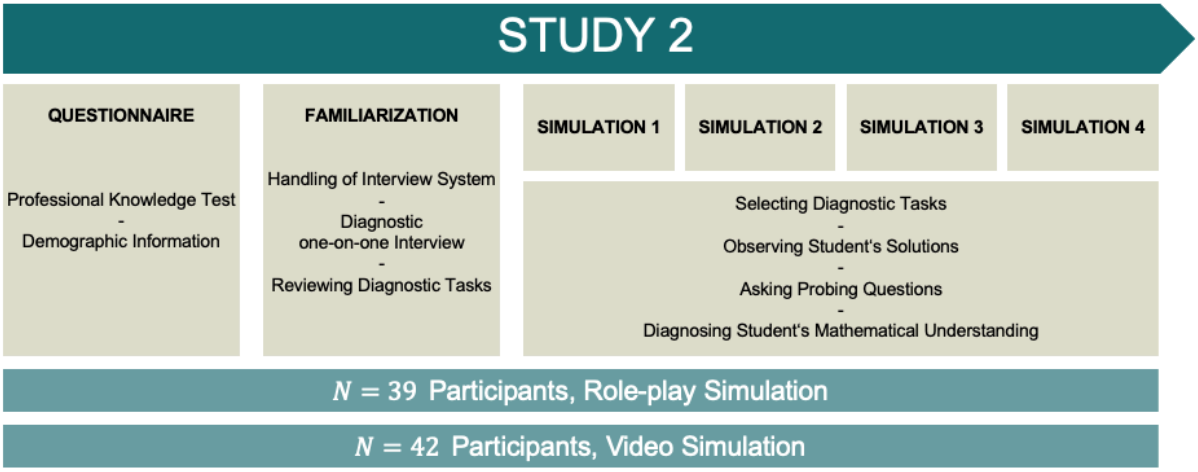

Figure 6: Design of the second study; findings reported in Manuscript C

4.5 Measures and Instruments

The first and second manuscripts (Manuscripts A and B) included in this dissertation analyzed the operationalization and relationship of different indicators of diagnostic competence (i.e., individuals' resources, diagnostic processes, and diagnostic performance) using data from Study 1. The third manuscript (Manuscript C) investigated the effects of the simulation's presentation format and the perception of the simulation on the development of individuals' interest as a key learning prerequisite, analyzing the data from Study 2 (see Figure 3).

4.5.1 Indicators of Diagnostic Competence

4.5.1.1 Pre-service Teachers' Resources

Regarding cognitive resources, the participants' professional knowledge was assessed. Following the categorization of Shulman (1987), a professional knowledge test comprising items assessing content knowledge (CK), pedagogical content knowledge (PCK), and pedagogical knowledge (PK) was constructed. The professional knowledge test was designed to cover the knowledge needed to accomplish the simulated diagnostic situation. The CK test refers to mathematical content knowledge of decimal fractions consisting of 12 items. For PCK, eight items addressing the teaching and learning of decimal fractions were constructed. To assess participants' PK, 11 items of the referring scale of the KiL project (Kleickmann et al., 2014) covering general aspects of assessment from a psychological perspective were adopted. All answers were coded dichotomously. By using a logistic Rasch model

(Rasch, 1960), individual person-parameters for each knowledge component (CK, PCK, PK) were calculated[10].

As an affective-motivational resource, participants' interest in diagnosis and interest in mathematics education were assessed using adaptations of the established scales of Rotgans and Schmidt (2011). The scale consisted of three items answered on a five-point Likert scale. Individual interest scores for interest in diagnosis and interest in mathematics education were calculated.

4.5.1.2 Diagnostic Process

The participants' selection of diagnostic tasks was assessed as an indicator of the diagnostic process. The diagnostic potential of the single tasks was not explicitly labeled in the simulations, so the participants needed to screen the task set in the familiarization phase to evaluate the diagnostic potential of each task. Each task from the given task set was coded dichotomously, having high or low diagnostic potential. Diagnostic processes drawing on more high potential tasks were assumed to reflect the higher quality of the diagnostic process.

4.5.1.3 Diagnostic Performance

After the diagnostic one-one-one interview, the participants were asked to rate the simulated student's mathematical understanding in nine subareas of decimal fractions (see Section 4.3). This rating was used to calculate the accuracy of the participants' diagnostic judgments. The judgment was assessed in a closed-answer format (response options: "student mastered", "student did not master", and "diagnosis not possible"). Answers of "diagnosis not possible" were coded with 0.5 points, considering abstention to be more helpful than risking an incorrect diagnosis. Indeed, this response option was very rarely chosen. Answers of "student mastered" or "student did not master" were coded with 1 point when correct and 0 points when incorrect. The achieved score was averaged over the nine subareas, resulting in an accuracy score as a performance indicator of the participant's diagnostic competence.

[10] For detailed information regarding the calculation of knowledge person-parameters and the used scaling sample, see Section 2.1.1 of Manuscript B.

4.5.2 Perception of the Simulation

As a measure of the perception of the simulation, established scales for authenticity, immersion, and extraneous cognitive load were used. Authenticity was assessed by three items (Seidel et al., 2010), immersion by four items (Frank, 2015), and extraneous cognitive load by three items (Opfermann, 2008), each on a five-point Likert scale.

5 Summary of Manuscripts

5.1 Manuscript A: *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*

Considering criticism regarding the absence of research on the characteristics of the diagnostic process, the first manuscript aimed to operationalize the characteristics of the diagnostic process as an indicator of diagnostic competence, in addition to individuals' resources and diagnostic performance. By focusing on the conceptualization of process measures, this manuscript broadens the research field of diagnostic competence, which has commonly addressed performance measures of diagnostic competence, most prominently judgment accuracy. Analyzing the data from Study 1, the aim of this manuscript was to contribute to the understanding of diagnostic competence by investigating interindividual differences during the diagnostic process and, more precisely, by investigating pre-service teachers' selection of diagnostic tasks for diagnostic one-on-one interviews in relation to their professional knowledge as cognitive resources. Based on theoretical assumptions, relations between all three components of professional knowledge (CK, PCK, and PK) and the selection of tasks could have been expected, but empirical evidence about such relations has been missing so far (Tröbst et al., 2018).

The taskset from which the participants could choose during the diagnostic one-on-one interviews contained tasks with high and low diagnostic potential. The selection of high-potential tasks was considered a process indicator of diagnostic competence. Two aspects regarding the selection of diagnostic tasks were considered. With *sensitivity to the diagnostic task potential*, the higher odds of selecting tasks with high diagnostic potential were termed. The other aspect, *adaptive use of diagnostic task potential*, was used to indicate participants' task selection considering the task's diagnostic potential in combination with information about a student's understanding that could have been extracted from prior tasks.

The analyses of participants' task selection using generalized linear mixed models (Bates et al., 2014) to take the dataset's nested structure into account showed

interindividual differences for the sensitivity to diagnostic task potential but not for the adaptive use of tasks' diagnostic potential. Regarding participants' professional knowledge, only CK had a significant effect, but not PCK, indicating that the discard of low-potential tasks is primarily affected by content-focused, mathematical analyses. Moreover, the results indicate that pre-service teachers might benefit from assistance in becoming aware of, and making use of, diagnostic task potential.[11]

The manuscript has been published as open access in the international journal *Frontiers in Education* as part of the research topic "Exploring Classroom Assessment Practices and Teacher Decision Making*".* The peer review process was double-blind.

The design of the study was developed by all the authors. Stefan Ufer initiated the project and supported the design of the simulation, data collection, and data analysis. He revised the analyses and manuscript. Daniel Sommerhoff supported, during the design of the simulation, the creation of the scales for professional knowledge and data analysis and revised the analysis scripts and the manuscript. Maike Achtner supported the data collection. The first author (Stephanie Kron) supervised the data collection, analyzed the data, and wrote the first draft of the manuscript. She incorporated the coauthors' and reviewers' remarks. She was responsible for the submission process and serves as the corresponding author for this manuscript.

5.2 Manuscript B: *Cognitive and Motivational Person Characteristics as Predictors of Diagnostic Performance: Combined Effects on Pre-Service Teachers' Diagnostic Task Selection and Accuracy*

In light of the results of Manuscript A, which did not find an effect of participants' PCK on the selection of tasks with high diagnostic potential and the theoretical assumptions about the moderating role of affective-motivational resources on the effects of cognitive resources (Herppich et al., 2018), we expected affective-motivational resources to be crucial for the application of knowledge during simulations of diagnostic situations. The moderating effect of affective-motivational resources is assumed by Herppich et al. (2018), but empirical investigations are scarce. The aim of this manuscript was to propose a concrete mechanism for the effect of affective-motivational resources on the

[11] See Appendix, Manuscript A.

effects of cognitive resources (i.e., individuals' interest facilitating knowledge activation and use), contributing to the understanding of the interplay of individuals' resources.[12] More precisely, combined effects of cognitive and affective-motivational resources and their interaction terms on the process and performance indicators of diagnostic competence had been investigated, again using data from Study 1. As cognitive resources, again, the participants' CK, PCK, and PK were considered. Regarding affective-motivational resources, participants' interest in diagnosis and interest in mathematics education were observed. The ratio of high-potential tasks, reflecting the participants' sensitivity to diagnostic task potential, was again used as a process indicator of diagnostic competence. To also consider a performance indicator of diagnostic competence, the accuracy of the diagnostic judgment was investigated.

Linear mixed models (Bates et al., 2014), accounting for the dataset's nested structure, were used to investigate the main and interaction effects of the participants' professional knowledge components and interest facets on process and performance indicators.

Regarding the main effects on the characteristics of the diagnostic process, only the effect of CK reached significance, in accordance with the findings of Manuscript A. For the quality of diagnostic performance, also only the effect of CK reached significance. Including interaction terms between single professional knowledge components and interest facets extended the findings. Significant interactions of CK and interest in diagnosis on accuracy, PCK and interest in mathematics education on accuracy and on the ratio of high-potential tasks, PK and interest in diagnosis on accuracy, and PK and interest in mathematics education on the ratio of high-potential tasks, support the assumption that affective-motivational resources are relevant for the activation and use of knowledge, at least in terms of simulation-based settings.

Seeing the relevance of both cognitive and affective-motivational resources, simulation-based learning environments should either support less interested participants in their activation of professional knowledge or should be designed to arouse participants' interest.[13]

[12] The term "person characteristics", used in Manuscript B, is to be used synonymously for the term "resources".
[13] See Appendix, Manuscript B.

The manuscript is published open access in the English language in a German journal having an international audience, *Journal für Mathematik-Didaktik,* as part of the research topic "Diagnostic thinking and activities of mathematics teachers—Theoretical foundations, empirical findings, and implications for practice" (original title: "Diagnostisches Denken und Handeln von Mathematiklehrkräften—Theoretische Fundierungen, empirische Befunde und Implikationen für die Praxis"). The peer review process was single-blind.

The design of the study was developed by all the authors. Stefan Ufer initiated the project and supported the design of the simulation, data collection, and data analysis. He revised the analyses and manuscript. Daniel Sommerhoff supported the design of the simulation and data analysis and revised the analysis scripts and manuscript. Kathleen Stürmer, Christof Wecker, and Matthias Siebeck supported the creation of the employed scales. Maike Achtner supported the data collection. The first author (Stephanie Kron) supervised the data collection, analyzed the data, and wrote the first draft of the manuscript. She incorporated the coauthors' and reviewers' remarks. She was responsible for the submission process and serves as the corresponding author for this manuscript.

5.3 Manuscript C: *Simulation-based Learning Environments: Do They Affect Learners' Relevant Interests?*

The third publication of this dissertation takes up one of the implications derived from Manuscript B, seeing individuals' interest as a facilitator of knowledge activation and use, as well as theoretical assumptions proposed by Hidi and Renninger (2006) about positive effects of authentic learning experiences on the development of interest, investigating whether learning experiences made in an authentic and immersive situation like simulated diagnostic interviews can affect individuals' interests as resources underlying diagnostic competence. Aiming to gain insights with respect to the different presentation formats used for the simulations, the data from Study 2 were investigated. In this study, the participants conducted four role-play or four video simulations. The perception of the simulation, conceptualized by perceived authenticity, immersion, and extraneous cognitive load, was assessed during each simulation. Again, the participants' interest in diagnosis and mathematics education

was considered. Interest facets were assessed before the first simulation and at the end of each single simulation.

Again, due to the nested structure of the dataset, linear mixed models (Bates et al., 2014) were used to estimate the effects of the presentation format of the simulation and the perception of the simulation on the development of participants' interest.[14]

It revealed that higher perceptions of authenticity and immersion went along with higher levels of interest, whereas the perception of extraneous cognitive load showed a negative relation to the participants' interest. The results showed that the development of the participants' interests was only weakly related to the simulation's presentation format when controlling for the perception of the simulation. Thus, the use of explicit interventions embedded in the simulation-based learning environment to develop and sustain learners' interests could be beneficial.

The manuscript has been published open access in the conference proceedings of the *45th Conference of the International Group for the Psychology of Mathematics Education.* The peer review process was double-blind.

The design of the study was developed by all the authors. Stefan Ufer initiated the project and supported the design of the simulation, data collection, and data analysis. He revised the analyses and manuscript. Daniel Sommerhoff supported, during the design of the simulation, the creation of the scales for professional knowledge and data analysis and revised the analysis scripts and the manuscript. Kathleen Stürmer, Christof Wecker, and Matthias Siebeck supported the creation of the employed scales. Maike Achtner supported the data collection. The first author (Stephanie Kron) supervised the data collection, analyzed the data, and wrote the first draft of the manuscript. She incorporated the coauthors' remarks. She was responsible for the submission process and serves as the corresponding author for this manuscript.

[14] See Appendix, Manuscript C.

6 Discussion

The three manuscripts included in this dissertation contribute to the understanding of diagnostic competence by investigating individuals' resources, characteristics of the diagnostic process, the quality of diagnostic performance as indicators of diagnostic competence, and the relations of these indicators. To address the call for a close to real-life assessment of competence, a simulation of diagnostic one-on-one interviews was established in two different presentation formats (i.e., role-play and video simulation). The research findings and the implications derived are discussed in subsequent sections, following the research desiderata raised in the background section.

6.1 Operationalization of Task Selection as a Process Indicator of Diagnostic Competence

Aiming to broaden the research on teachers' diagnostic competence by investigating the characteristics of the diagnostic process as indicators of diagnostic competence, this dissertation investigated mathematics pre-service teachers' task selection during diagnostic interviews. The aim of this endeavor was to gain insights into relevant process indicators of diagnostic competence. The selection of tasks with high diagnostic potential was operationalized as sensitivity to the diagnostic potential of the tasks. A real-time adaptation of the selection of high-potential tasks, done in accordance with the diagnostic evidence that could have been extracted from prior tasks about the specific student, was operationalized as an adaptive use of diagnostic task potential.

The results revealed that the participants did not show a higher probability of selecting high-potential tasks. However, significant interindividual variation between participants in their tendency to prefer high-potential tasks over low-potential tasks was observed, indicating that the selection of high-potential tasks indeed reflects an important characteristic of the diagnostic process and, as such, is seen as a promising process indicator of diagnostic competence. From the observation of interindividual differences, we derive the implication that at least those participants who were not able to make a sensitive use of the diagnostic potential of tasks may benefit from support in attending to task potential. The results of this dissertation suggest that identifying and making

use of a task's diagnostic potential is related to pre-service teachers' CK. Thus, further research should investigate whether prompts supporting pre-service teachers in activating their CK, or prompts offering the CK necessary to accomplish the simulated task, relate to an increase in the quality of pre-service teachers' diagnostic processes. The significant interaction effect between pre-service teachers' interest in mathematics education and their PCK on the selection of high-potential tasks calls for specific support. Again, prompts activating PCK, as well as prompts providing the necessary PCK, could be promising (Irmer et al., 2022). Moreover, interventions focusing on the development of PCK could be an effective approach, as shown by Ostermann et al. (2018), concerning pre-service teachers' ability to estimate task difficulties.

For the construct of an adaptive use of diagnostic task potential, no significant effects or interindividual variation could be observed. Because adaptive use requires a combination of knowledge about a task's diagnostic potential and the real-time interpretation and integration of student responses to prior tasks, different explanations for why effects may have been vanished are possible, such as interindividual differences in combining knowledge about diagnostic task potential and student understanding or insufficient knowledge about diagnostic task potential or the student's understanding. Measuring adaptive use, as it was operationalized in this dissertation, is methodically intricate, since the participants' actual interpretation and integration of student responses had not been accessed explicitly.

6.2 Relation of Individuals' Resources to Process and Performance Indicators of Diagnostic Competence

The construct of diagnostic competence was investigated by analyzing the process and performance indicators of diagnostic competence and their relationship to individuals' resources. The aim was to gain insights about which individual resources are needed to accomplish professional situations and whether those resources interact with each other to derive implications whereby (e.g., task selection, knowledge activation) prompts could be used to support individuals and how they should be designed (e.g., providing relevant CK or PCK).

Regarding the main effects of professional knowledge on process and performance indicators, it was only pre-service teachers' CK, but not their PCK, which showed a significant relationship to process and performance indicators, indicating that the

selection of tasks was based on analyzing the tasks' mathematical content, and that the diagnostic judgment rather referred to evaluating the mathematical correctness of student solutions. Since knowledge about diagnostic task potential and student misconceptions is discussed as a facet of PCK (Baumert & Kunter, 2013), primarily relations between PCK and the indicators of diagnostic competence had been expected. Even though Ostermann et al. (2018) did observe a significant relation between pre-service teachers' PCK and their diagnostic accuracy, van den Kieboom et al. (2014) likewise only observed a significant relation between pre-service teachers' CK and their questioning behavior in diagnostic interviews. When discussing these results, two possible explanations are available. First, not observing the main effects of PCK may be due to the fact that the individuals in our sample have not acquired the necessary PCK knowledge yet, which at least applies to some participants in our sample, as indicated by their knowledge scores on our professional knowledge test. Those participants could benefit from explicit interventions focusing on the development of PCK, as was done by Tröbst et al. (2019). On the other hand, those participants who seemed to have acquired the respective PCK, as indicated by the knowledge scores gained in the professional knowledge test, may have struggled to activate and apply this knowledge in the simulated setting. These participants may benefit from scaffolding that focuses on knowledge activation and repeated participation in simulation-based learning environments to train them in the application of acquired knowledge. Moreover, affective-motivational resources may impact the activation and application of knowledge; therefore, scaffolds focusing on participants' affective-motivational resources may be promising, as well (Nickl et al., 2023).

None of the investigated main effects of participants' interest (i.e., in mathematics education, interest in diagnosis) reached significance. From a methodological perspective, this could rely on the fact that a significant main effect of interest would only be plausible if enough interested participants would also have sufficient professional knowledge to accomplish the simulated task. According to the results of the professional knowledge test, a substantial number of participants did not have sufficient knowledge. This means that even if they had been highly engaged due to their high interest, they would not have been able to meet the professional demands of the simulated situation because of a lack of professional knowledge.

We derive from these results that interest or professional knowledge alone can explain only a few differences in individuals' diagnostic processes and diagnostic performance, initiating an investigation into the interaction effects of those two individual resources.

The results of the respective analyses showed that including the interaction effects of interest facets on the relationship between professional knowledge components and process or performance indicators significantly improved the explanation for individual differences in the investigated process or performance indicator. The results of this analyses provide empirical evidence for the assumption proposed by Herppich et al. (2018) that affective-motivational resources moderate the effects of cognitive resources. The results indicate that individuals' interests seem to promote the identification and activation of knowledge relevant to a specific learning situation (Hidi, 1990).

The findings indicate that interest seems to act as a facilitator of knowledge activation and use. In the context of competence assessment and development, this implies that it is not necessarily sufficient for an individual to acquire the respective cognitive resources. Rather, it is also necessary that the corresponding affective-motivational resources (e.g., interest) are sufficiently present to support the activation of acquired knowledge. Scaffolds aiming to arouse participants' interest to support the activation of knowledge, or scaffolds supporting less interested individuals in their activation of knowledge, could be promising in this vein (Nickl et al., 2023).

6.3 The Role of Presentation Format and Perception of a Simulation on the Development of Individuals' Interest

Simulations are a promising approach to fostering professional competences in higher education (Chernikova et al., 2020b). Building on the insights derived from the first two manuscripts included in this dissertation regarding the relationships between individuals' resources and the different indicators of diagnostic competence assessed in a simulation, we also investigated whether individuals' resources (interest) were affected by the characteristics of the simulation itself. Considering the role of individuals' interest in the activation of knowledge and, as such, in process and performance indicators, the development of interest in relation to the presentation format and the perception of the simulation was investigated. Following Hidi and Renninger (2006), experiences made during authentic and immersing learning

situations can trigger situational interest, which may contribute to the development of individual interest. On the other hand, learning situations causing a high cognitive load have been found to negatively affect situational interest (Park et al., 2015). Whereas situational interest may enhance learning (Wade, 1992), individual interest has been found to positively affect attention, recognition, and recall (Hidi & Renninger, 2006). To investigate the potential effects of the learning experiences made in a simulation-based learning environment, the participants' situational interest after each of the four simulations was assessed. One must consider that it is hardly possible to derive assumptions about the causal relationship between the perception of the simulation and the level of situational interest since it was assessed during the simulation. Confirming the assumptions of Hidi and Renninger (2006), participants who perceived the simulation as authentic and immersive reported higher levels of interest. The negative relationship between extraneous cognitive load and the development of interest is in line with the findings of Park et al. (2015). This underpins the call of Codreanu et al. (2020) to find a balance between authenticity and cognitive demand when designing AoPs.

In terms of functional correspondence, the role-play simulation may have had a higher authenticity (e.g., freely asking verbal probing questions) than the video simulation, in which pre-defined probing questions were displayed. With respect to physical resemblance, the video simulation may have appeared quite authentic as the participants heard the voice of a real 6th grader, amended by a picture of a 6th grader (Chernikova et al., 2023), whereas the role of the 6th grader in the role-play simulation was played by adult trained research assistants, which may have lowered the physical resemblance of the role-play simulation. Regarding cognitive load, the video simulation may have caused a lower extraneous cognitive load, since pre-defined probing questions were displayed by the interview system used for the diagnostic interviews and the participants did not have to interact directly with a student (Sweller, 2010).

Beyond the perception of the simulation, the development of interest was only weakly related to the presentation format of the simulation. Based on the findings presented, it is hardly possible to derive reliable assumptions about the interdependency between the perception of the simulation or its presentation format and the development of interest. In terms of interest development, both presentation formats can be used likewise, until those mechanisms are grasped by empirical investigations.

6.4 Limitations and Further Research

The findings presented in this dissertation are constricted by different limitations that call for further research. First and foremost, the datasets analyzed in this dissertation do not contain any data on in-service teachers, which limits the generalizability of findings to pre-service teachers' diagnostic competence. However, it allows us to derive implications for challenges in the initial acquisition of these competence indicators. Further research is needed to investigate the transferability of the findings to in-service teachers to derive implications for in-service teacher training. Moreover, the sample sizes of $N = 65$, $N = 63$, and $N = 81$ pre-service teachers limit the statistical power of the presented studies, which could also explain the absence of significant effects (e.g., main effect of PCK), in turn calling for further investigations with a larger sample size, including in-service teachers.

The insights derived for the diagnostic competence of pre-service teachers refer to the diagnosis of student understanding in very specific content (i.e., decimal fractions), in which a set of potential student misconceptions to be diagnosed is given. Further research investigating the resilience of findings with respect to other content areas, or areas of student understanding in which systematizations of student misconceptions are missing, is needed (cf. judgment specificity, Südkamp et al., 2012).

Regarding the operationalization of process indicators of diagnostic competence, the presented approach is very specific to the simulated professional demand of selecting tasks for a diagnostic one-on-one interview, only approximating teachers' everyday working lives (e.g., selecting tasks for instruction). Since the selection of tasks is only one aspect of the diagnostic process, participants' asking of probing questions should also be analyzed, which in turn may be a promising process indicator as shown by van den Kieboom et al. (2014), investigating pre-service teachers' questions during diagnostic one-on-one interviews in the area of algebraic thinking. Further research investigating process indicators with respect to the professional vision framework or different diagnostic activities could be beneficial for the conceptualization of process indicators (Codreanu et al., 2021; Kramer et al., 2021; Seidel & Stürmer, 2014). Further research is needed to investigate potential explanations for not observing effects of an adaptive use of diagnostic potential. The question arises of how pre-service teachers integrate the information gained about the student's understanding into their diagnostic

processes. Alternative operationalizations of the adaptive use of diagnostic task potential, addressing this gap, are needed.

The investigation of relations between different indicators of diagnostic competence was limited due to the small sample sizes of the studies, and thus it can only be a first step toward an understanding of the interplay of the different indicators. Larger sample sizes would also allow the investigation of the effects of additional affective-motivational (state) variables or cognitive load (Codreanu et al., 2020), in order to explain the moderating effects of affective-motivational resources in more detail. Based on the findings of this dissertation, further research is needed to investigate the relationship between process and performance indicators. In addition, research on the joint explanatory power of individuals' resources and their processes for performance, as operationalized by Hammer and Ufer (2023) for teachers' dealing with tasks in lesson planning, is an open object for the future.

This dissertation focused on using AoPs (i.e., simulations) as an assessment tool to measure indicators of diagnostic competence close to real-life settings. Due to the COVID-19 pandemic, the data collection in Study 2 had to be done online, which may have affected the quality of the data. Moreover, the role-play simulation took place online instead of in an in-person meeting, as it was supposed to be. This circumstance may have affected the participants' levels of immersion and their perceptions of the simulation. Considering this, the results regarding the development of individuals' interest in accordance with the presentation format and the perception of the simulation are limited. Even though the dataset did not indicate a lack of quality (e.g., no missing values, similar levels of interaction between the participants and the simulated students, just as in the role-play simulations of the first study, which were done in person), repeating the data collection for this study in an in-person setting could be beneficial for a renewed analysis of the effects observed. On the other hand, observing a kind of "digital turn" in higher education in the wake of the COVID-19 pandemic, one must consider that video simulations prospectively may be more practicable than role-play simulations, which may need to be transferred to online settings as well. Because only the development of interest was investigated, further research is needed to analyze whether the presentation format or the perception of the simulation affects indicators of the diagnostic process, performance, or participants' cognitive resources.

7 Conclusion

Teachers' diagnostic competence is seen as a key prerequisite for adaptive and effective teaching. As such, pre- and in-service teacher trainings aim to facilitate teachers' diagnostic competence. In this regard, a promising approach is the use of simulation-based learning environments, which can also be used to assess diagnostic competence close to real-life situations in controlled circumstances. For the design of simulations to assess and foster teachers' diagnostic competence, it is necessary to understand what contributes to the complex construct of diagnostic competence. The research presented in this dissertation addresses this necessity by investigating the relationship between individuals' resources and indicators of the diagnostic process and performance in the context of simulated diagnostic one-on-one interviews.

This dissertation broadens the research field of teachers' diagnostic competence, commonly investigating performance indicators of diagnostic competence by proposing operationalizations of process indicators. The observation of significant inter-individual variation between participants in their tendency to select tasks with high diagnostic potential during the diagnostic interviews indicates that the selection of high-potential tasks indeed reflects an important process indicator of diagnostic competence. The investigation of the characteristics of the diagnostic process, in addition to the quality of the diagnostic performance mirrored in judgment accuracy, enriches the research field of diagnostic competence by addressing diagnostic competence holistically.

The investigation of the relationships between pre-service teachers' cognitive and affective-motivational resources, as well as their interactions, to indicators of the diagnostic process and indicators of diagnostic performance contributes to the research field by providing empirical evidence for the theoretical assumption of affective-motivational resources moderating the effects of cognitive resources. Significant interaction effects between participants' CK and PCK, and their interest in diagnosis and interest in mathematics education, on the selection of tasks having high diagnostic potential and the accuracy of the diagnostic judgment, imply that diagnostic competence requires the interplay of sufficient cognitive (e.g., professional knowledge) and affective-motivational resources (e.g., interest).

Taking up the important role of individuals' interest, its development in consideration of the presentation format of the simulation and the participants' perception of the simulation was investigated. It revealed that those participants who perceived the simulation as authentic and immersing reported higher levels of interest, whereas the perception of extraneous cognitive load was negatively related to participants' interest reported after having conducted a simulated diagnostic one-on-one interview.

The assessment of process and performance indicators close to real life in a simulated diagnostic one-on-one interview addresses the call for a competence assessment close to professional demands and contributes to the robustness of the derived implications. The significant interaction effects between participants' cognitive and affective-motivational resources on process and performance indicators of diagnostic competence provide empirical evidence for the theoretical assumptions of Herppich et al. (2018), underpinning the important role of both cognitive and affective-motivational resources. The insights of this dissertation are especially relevant for the design of simulation-based learning environments and scaffolds used to foster the development of pre-service teachers' diagnostic competence.

8 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

Albu, C., & Lindmeier, A. (2023). Performance assessment in teacher education research – A scoping review of characteristics of assessment instruments in the DACH region. *Zeitschrift für Erziehungswissenschaft*, *26*(3), 751-778. https://doi.org/10.1007/s11618-023-01167-7

Alibali, M. W., & Sidney, P. G. (2015). Variability in the natural number bias: Who, when, how, and why. *Learning and Instruction*, *37*, 56-61. https://doi.org/10.1016/j.learninstruc.2015.01.003

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Package Lme4: Linear Mixed-Effects Models Using Eigen and S4. *Journal of statistical software 67*, 1-103.

Baumert, J., & Kunter, M. (2013). The COACTIV Model of Teachers' Professional Competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results from the COACTIV Project* (pp. 25-48). Springer. https://doi.org/10.1007/978-1-4614-5149-5_2

Behrmann, L., & Souvignier, E. (2013). The Relation Between Teachers' Diagnostic Sensitivity, their Instructional Activities, and their Students' Achievement Gains in Reading. *Journal of Educational Psychology*, *27*(4), 283-293. https://doi.org/10.1024/1010-0652/a000112

Binder, K., Krauss, S., Hilbert, S., Brunner, M., Anders, Y., & Kunter, M. (2018). Diagnostic Skills of Mathematics Teachers in the COACTIV study. In T. Leuders, K. Philipp, & J. Leuders (Eds.), *Diagnostic Competence of Mathematics Teachers - Unpacking a complex construct in teacher education and teacher practice* (pp. 33-53). Springer. https://doi.org/10.1007/978-3-319-66327-2_2

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. (2015). Beyond dichotomies: Viewing competence as a continuum. *Zeitschrift für Psychologie [Journal of Psychology]*, *223*, 3-13. https://doi.org/10.1027/2151-2604/a000194

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., Fischer, F., & COSIMA, D. F. G. R. g. (2020a). 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. (2020b). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, *20*(10), 499-541. https://doi.org/10.3102/0034654320933544

Chernikova, O., Holzberger, D., Heitzmann, N., Stadler, M., Seidel, T., & Fischer, F. (2023). Where salience goes beyond authenticity. *Zeitschrift für pädagogische Psychologie*, 1-11. https://doi.org/https://doi.org/10.1024/1010-0652/a000357

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

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12-25. https://doi.org/https://doi.org/10.1016/j.tate.2013.03.001

Fischbein, E., Deri, M., Nello, M. S., & Marino, M. S. (1985). The Role of Implicit Models in Solving Verbal Problems in Multiplication and Division. *Journal for research in mathematics education*, *16*, 3-17. https://doi.org/10.5951/jresematheduc.16.1.0003

Fischer, F., & Opitz, A. (2022). *Learning to Diagnose with Simulations - Examples from Teacher Education and Medical Education* (Springer Briefs in Education ed.). Springer. https://doi.org/10.1007/978-3-030-89147-3

Frank, B. (2015). *Presence messen in laborbasierter Forschung mit Mikrowelten*. Springer. https://doi.org/10.1007/978-3-658-08148-5

Goeze, A., Zottmann, J. M., Vogel, F., Fischer, F., & Schrader, J. (2014). Getting immersed in teacher and student perspectives? Facilitating analytical competence using video cases in teacher education. *Instructional Science*, *42*(1), 91-114. https://doi.org/10.1007/s11251-013-9304-3

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. https://doi.org/10.1177/016146810911100905

Hammer, S., & Ufer, S. (2023). Professional competence of mathematics teachers in dealing with tasks in lesson planning. *Teaching and Teacher Education*, *132*, 1-13. https://doi.org/https://doi.org/10.1016/j.tate.2023.104246

Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic medicine*, *89*(3), 387-392. https://doi.org/10.1097/ACM.0000000000000130

Heckmann, K. (2006). Zum Dezimalbruchverständnis von Schülerinnen und Schülern: Theoretische Analyse und empirische Befunde. Berlin: Logos.

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M., Ufer, S., Schmidmaier, R., Neuhaus, B., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., & Fischer, F. (2019). Facilitating diagnostic competences in simulations: A conceptual framework and a research agenda for medical and teacher education. *FRONTLINE LEARNING RESEARCH*, *7*, 1-24. https://doi.org/10.14786/flr.v7i4.384

Helmke, A. (2009). Unterrichtsqualität und Lehrerprofessionalität Diagnose, Evaluation und Verbesserung des Unterrichts [Instructional quality and teacher professionalism. Diagnosis, evaluation and improvement of instruction]. Seelze-Velber: Klett/Kallmeyer.

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-Peters, 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. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*(4), 549-571. https://doi.org/10.3102/00346543060004549

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

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*, *44*(9), 1506-1526. https://doi.org/10.1080/09500693.2022.2083253

Kaiser, G., & König, J. (2019). Competence Measurement in (Mathematics) Teacher Education and Beyond: Implications for Policy. *Higher Education Policy*, *32*(4), 597-615. https://doi.org/10.1057/s41307-019-00139-z

Kleickmann, T., Großschedl, J., Harms, U., Heinze, A., Herzog, S., Hohenstein, F., Köller, O., Kröger, J., Lindmeier, A., & Loch, C. (2014). Professionswissen von Lehramtsstudierenden der mathematisch-naturwissenschaftlichen Fächer - Testentwicklung im Rahmen des Projekts KiL. *Unterrichtswissenschaft - Zeitschrift für Lernforschung*, *42*(3), 280-288.

Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms bei der DFG [Competence models for assessing individual learning outcomes and evaluating educational processes. Description of a new priority program of the German Research Foundation, DFG]. *Zeitschrift für Pädagogik*, *52*(6), 876-903. https://doi.org/10.25656/01:4493

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

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*, 1-24. 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 Journal*, *7*, 1-28. https://doi.org/10.1080/2331186X.2020.1823155

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*(4), 383-409. https://doi.org/10.1016/s0959-4752(01)00011-1

Kron, S., Sommerhoff, D., Stürmer, K., & Ufer, S. (submitted). Design and implementation of a video simulation to assess and foster pre-service teachers' diagnostic competences. In F. Schultz-Pernice, S. Berger, & I. Mausz (Eds.), *Das Beste aus beiden Welten*. Springer.

Leuders, T., Dörfler, T., Leuders, J., & Philipp, K. (2018). Diagnostic Competence of Mathematics Teachers: Unpacking a Complex Construct. In T. Leuders, T. Dörfler, J. Leuders, & K. Philipp (Eds.), *Diagnostic Competence of Mathematics Teachers. Unpacking a Complex Construct in Teacher Education and Teacher Practice* (pp. 3-32). Springer. https://doi.org/10.1007/978-3-319-66327-2_1

Loibl, K., Leuders, T., & Dörfler, T. (2020). A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teaching and Teacher Education*, *91*, 103059. https://doi.org/10.1016/j.tate.2020.103059

Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why Is Learning Fraction and Decimal Arithmetic So Difficult? *Developmental Review*, *38*, 201-221. https://doi.org/10.1016/j.dr.2015.07.008

Marczynski, B., Kaltefleiter, L. J., Siebeck, M., Wecker, C., Stürmer, K., & Ufer, S. (2022). Diagnosing 6th graders´ understanding of decimal fractions. Fostering mathematics pre-service teachers' diagnostic competences with simulated one-to-one interviews. In F. Fischer & A. Opitz (Eds.), *Learning to diagnose with simulations - examples from teacher education and medical education* (pp. 49-62). Springer. https://doi.org/10.1007/978-3-030-89147-3_5

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741-749. https://doi.org/10.1037/0003-066x.50.9.741

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*(1), 45. https://doi.org/10.1186/s41239-022-00351-9

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*, 1-8. https://doi.org/https://doi.org/10.1024/1010-0652/a000362

Opfermann, M. (2008). There's more to it than instructional design: The role of individual learner characteristics for hypermedia learning. Berlin: Logos.

Ostermann, A. (2018). Factors Influencing the Accuracy of Diagnostic Judgments. In T. Leuders, K. Philipp, & J. Leuders (Eds.), *Diagnostic Competence of Mathematics Teachers - Unpacking a complex construct in teacher education and teacher practice* (pp. 95-108). Springer. https://doi.org/10.1007/978-3-319-66327-2_5

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

Padberg, F., & Wartha, S. (2017). *Didaktik der Bruchrechnung*. Berlin: Springer Spektrum.

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior*, *44*, 267-278. https://doi.org/10.1016/j.chb.2014.10.061

Rach, S. (2021). Relations between individual interest, experiences in learning situations and situational interest. In M. Inprasitha, N. Changsri, & N. Boonsena (Eds.), *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 491-499). PME.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Resnick, L., Nesher, P., Léonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual Bases of Arithmetic Errors: The Case of Decimal Fractions. *Journal for research in mathematics education*, *20*(1), 8-27. https://doi.org/10.2307/749095

Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest in academic achievement in the active-learning classroom. *Learning and Instruction*, *21*(1), 58-67. https://doi.org/10.1016/j.learninstruc.2009.11.001

Schack, E., Fisher, M., Thomas, J., Eisenhardt, S., Tassell, J., & Yoder, M. (2013). Prospective elementary school teachers' professional noticing of children's early numeracy. *Journal of Mathematics Teacher Education*, *16*(5), 379-397. https://doi.org/10.1007/s10857-013-9240-9

Seidel, T., Blomberg, G., & Stürmer, K. (2010). Observer: Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. *Zeitschrift für Pädagogik*, *56*, 296-306. https://doi.org/10.25656/01:3438

Seidel, T., & Stürmer, K. (2014). Modeling and Measuring the Structure of Professional Vision in Preservice Teachers. *American educational research journal*, *51*, 739-771. https://doi.org/10.3102/0002831214531321

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*, 259-267. https://doi.org/10.1016/j.tate.2010.08.009

Shavelson, R. (2012). Assessing business-planning competence using the Collegiate Learning Assessment as a prototype. *Empirical Research in Vocational Education and Training*, *4*(1), 77-90. https://doi.org/10.1007/bf03546509

Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, *2*(1), 41-63. https://doi.org/10.1007/BF03546488

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1-23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Siegler, R. S., & Lortie-Forgues, H. (2015). Conceptual knowledge of fraction arithmetic. *Journal of Educational Psychology*, *107*(3), 909-918. https://doi.org/10.1037/edu0000025

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*, *83*, 1-13. https://doi.org/10.1016/j.learninstruc.2022.101689

Stegmann, K., Pilz, F., Siebeck, M., & Fischer, F. (2012). Vicarious learning during simulations: is it more effective than hands-on training? *Medical Education 46*(10), 1001-1008. https://doi.org/10.1111/j.1365-2923.2012.04344.x

Steinle, V., & Pierce, R. (2006). Incomplete or incorrect understanding of decimals: an important deficit for studnet nurses. In J. Novotná, H. Moraová, M. Krátká, & N. Stehlíková (Eds.), *Proceedings 30th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 161-168). PME.

Steinle, V., & Stacey, K. (2004). A Longitudinal Study of Students 'Understanding of Decimal Notation: An Overview and Refined Results. In *Proceedings of the 27th annual conference of the Mathematics Education Research Group of Australasia* (pp. 541-548).

Stürmer, K., Fütterer, T., Kron, S., Sommerhoff, D., & Ufer, S. (under review). What Makes a Simulation-Based Learning Environment for Pre-Service Teachers Authentic? The Role of Individual Learning Characteristics and Context-Related Features. *European Journal of Psychology of Education*.

Stürmer, K., Marczynski, B., Wecker, C., Siebeck, M., & Ufer, S. (2021). Praxisnahe Lerngelegenheiten in der Lehrerausbildung - Validierung der simulationsbasierten Lernumgebung DiMaL zur Förderung diagnostischer Kompetenz von angehenden Mathematiklehrpersonen. In N. Beck, T. Bohl, & S. Meissner (Eds.), *Vielfältig herausgefordert. Forschungs- und Entwicklungsfelder der Lehrerbildung auf dem Prüfstand* (pp. 57-69). Tübingen University Press. https://tobias-lib.ub.uni-tuebingen.de/xmlui/bitstream/handle/10900/111260/TuSE02_vielfaltig_02-02_stuermer.pdf?sequence=1&isAllowed=y

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://psycnet.apa.org/doi/10.1037/a0027627

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, *22*(2), 123-138. https://doi.org/10.1007/s10648-010-9128-5

Tröbst, S., Kleickmann, T., Depaepe, F., Heinze, A., & Kunter, M. (2019). Effects of Instruction on Pedagogical Content Knowledge About Fractions in Sixth-Grade Mathematics to Content Knowledge and Pedagogical Knowledge. *Unterrichtswissenschaft*, *47*, 79-97. https://doi.org/10.1007/s42010-019-00041-y

Tröbst, S., Kleickmann, T., Heinze, A., Bernholt, A., Rink, R., & Kunter, M. (2018). Teacher Knowledge Experiment: Testing Mechanisms Underlying the Formation of Preservice Elementary School Teachers' Pedagogical Content Knowledge Concerning Fractions and Fractional Arithmetic. *Journal of Educational Psychology*, *110*(8), 1049-1065. https://doi.org/10.1037/edu0000260

Ufer, S., & Neumann, K. (2018). Measuring competencies. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 433-443). Routledge. https://doi.org/10.4324/9781315617572

van den Kieboom, L. A., Magiera, M. T., & Moyer, J. C. (2014). Exploring the relationship between K-8 prospective teachers' algebraic thinking proficiency and the questions

they pose during diagnostic algebraic thinking interviews. *Journal of Mathematics Teacher Education*, *17*, 429-461. https://doi.org/10.1007/s10857-013-9264-1

Wade, S. E. (1992). How interest affects learning from text. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role ofinterest in learning and development* (pp. 255-277). Lawrence Erlbaum Associates, Inc.

Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen. In *Leistungsmessungen in Schulen* (pp. 17-32). Beltz.

Appendix

Note
For copyright reasons, appendices will not be included in the online publication of the dissertation.