TUM School of Social Sciences and Technology
Technische Universität München

Analyzing and fostering mathematics teachers' assessment competencies with digital simulations

Christian Schons

Vollständiger Abdruck der von der TUM School of Social Sciences and Technology der Technischen Universität München zur Erlangung eines **Doktors der Naturwissenschaften (Dr. rer. nat.)** genehmigten Dissertation.

Vorsitz: Prof. Dr. Claudia Nerdel

Prüfende der Dissertation:
1. Prof. Dr. Andreas Obersteiner
2. Prof. Dr. Kristina Reiss
3. Prof. Dr. Daniel Sommerhoff

Die Dissertation wurde am 07.11.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Social Sciences and Technology am 17.02.2025 angenommen.

Acknowledgements

Wir sollen heiter Raum um Raum durchschreiten,
An keinem wie an einer Heimat hängen,
Der Weltgeist will nicht fesseln uns und engen,
Er will uns Stuf' um Stufe heben, weiten.
Kaum sind wir heimisch einem Lebenskreise
Und traulich eingewohnt, so droht Erschlaffen;
Nur wer bereit zu Aufbruch ist und Reise,
Mag lähmender Gewöhnung sich entraffen.

Ausschnitt aus *Stufen* von Hermann Hesse (1877–1962)

Auf meiner Reise zur Promotion durfte ich großartige Menschen an weit verstreuten Orten kennenlernen und dafür bin ich sehr dankbar. In erster Linie danke ich dafür herzlich meinen beiden Betreuern Kristina Reiss und Andreas Obersteiner, die mir diese Reise mit sehr viele Möglichkeiten zur wissenschaftlichen und persönlichen Weiterentwicklung eröffnet haben und mir die Freiheit gelassen haben, dabei eigene Wege zu beschreiten.

Bedanken möchte ich mich außerdem bei meinem Mentor Frank Reinhold für die Unterstützung auf diesem Weg und die vielen Diskussionen, aus denen ich jedes Mal mit einigen Antworten und vielen neuen Fragen und Impulsen herausging. Bei Daniel Sommerhoff und Claudia Nerdel bedanke ich mich für ihre Bereitschaft, als Prüfer bzw. Leiterin der Prüfungskommission bereitzustehen.

Herzlich danken möchte ich auch den Kolleginnen und Kollegen von den Lehrstühlen für Didaktik der Mathematik der LMU und der TUM. Ich konnte in ganz unterschiedlichen Konstellationen mit Euch zusammenarbeiten und in allen Konstellationen hatten wir eine unterstützende und produktive Zusammenarbeit. Außerdem möchte ich mich bei allen Forscherinnen und Forschern aus dem COSIMA Projekt für die vielen produktiven Zusammenkünfte bedanken. Hier gilt mein Dank besonders Frank Fischer für die Koautorenrolle bei beiden Publikationen und die Begleitung meiner Promotion in „unserem TP1“. Für die Korrektur des Manuskripts danke ich besonders meinem Bürokollegen Michael D'Erchie, sowie Alina Kadluba und Philipp Kather. Bei Anselm Strohmaier bedanke ich mich für die wissenschaftliche und moralische Unterstützung sowie die vielen Erlebnisse und Erkenntnisse auf gemeinsamen Konferenzreisen.

I would also like to thank all the colleagues that welcomed me so warmly in Loughborough and Fredericton, especially Megan Foulkes, Hanna Weiers, and David Wagner.

Schließlich gilt mein größter Dank meiner Familie für ihre bedingungslose Unterstützung und Liebe. Danke Barbora und Mathilda, dass ihr da seid und wir unsere Herausforderungen immer gemeinsam meistern.

Abstract

Mathematics teachers need to master core practices to be prepared for classroom situations. Teachers' core practices include the assessment of students' mathematical competencies and misconceptions. Evidence suggests that the accuracy of teacher assessments is linked to student achievement in the classroom. While teachers often struggle to make accurate assessments of student competencies, little is known about how to effectively foster teachers' competencies in assessing students' mathematical competencies and misconceptions. Currently, there are few opportunities for evidence-based and practice-oriented training of these core practices during university education, making the development and evaluation of practice opportunities a relevant topic for mathematics educational research.

The overarching goal of this dissertation was to develop effective support for pre-service mathematics teachers' assessment competencies. To this end, we used a comprehensive competency framework to investigate the mechanisms underlying accurate assessments. A simulation was used for providing a practice-oriented and controlled situation to analyze and foster pre-service teachers' assessment competencies. Simulations have been used effectively in education, but represent a rather novel approach to investigate mathematics teachers' professional competencies. Different types of scaffolding were integrated into the simulation that were effective in fostering complex skills. The log data from the simulation were used to analyze pre-service teachers' assessment processes. This method provided an unobtrusive way of capturing the assessment process in an interactive situation, including the processes of task selection and the interpretation of student solutions.

This dissertation presents two publications that contribute to the development of effective support of pre-service teachers' assessment competencies. The first publication evaluated the design and effectiveness of a digital simulation with different types of scaffolding to enhance pre-service teachers' assessment accuracy. The second publication analyzed relations between pre-service teachers' dispositions, the assessment process, and assessment accuracy to inform the implementation of adaptive scaffolding.

The results indicate that simulations are suitable for analyzing and fostering pre-service mathematics teachers' assessment competencies. Scaffolding that focused on the pedagogical content was more effective in enhancing accuracy than scaffolding that focused on strategies in the assessment process. During the assessment process, pronounced individual differences were identified that can be used to develop adaptive scaffolding. Overall, this dissertation project can serve as an example for the development and evaluation of simulations aimed at fostering teachers' core practices during their university education.

Zusammenfassung

Mathematiklehrkräfte müssen professionelle Kompetenzen erlernen, um auf die Unterrichtspraxis vorbereitet zu sein. Zu diesen Kompetenzen gehört die Diagnose mathematischer Fähigkeiten und Fehlvorstellungen von Schülerinnen und Schülern. Studien zeigen, dass die Genauigkeit von Diagnosen mit dem Lernerfolg von Schülerinnen und Schüler im Unterricht zusammenhängt. Während es Lehrkräften häufig schwerfällt, mathematische Fähigkeiten präzise zu diagnostizieren, ist wenig darüber bekannt, wie die Diagnosekompetenzen von Lehrkräften effektiv gefördert werden können, um mathematische Fähigkeiten und Fehlvorstellungen zu diagnostizieren. Derzeit gibt es nur wenige evidenzbasierte und praxisnahe Möglichkeiten, Diagnoseompetenzen während der universitären Ausbildung zu trainieren, was die Entwicklung und Evaluation solcher praxisnahen Trainingsmöglichkeiten zu einem relevanten Forschungsthema der Mathematikdidaktik macht.

Das übergeordnete Ziel dieses Dissertationsprojekts war es, wirksame Unterstützungsmaßnahmen zur Förderung der Diagnosekompetenzen von angehenden Mathematiklehrkräften zu entwickeln. Hierfür wurde ein ganzheitliches Modell für Diagnosekompetenzen von Lehrkräften verwendet, um die Mechanismen zu untersuchen, die akkuraten Diagnosen zugrunde liegen. Die Verwendung einer Simulation ermöglichte eine praxisorientierte und gleichzeitig kontrollierte Situation, um die Diagnosekompetenzen angehender Lehrkräfte zu analysieren und zu fördern. Simulationen wurden zur Vorbereitung auf die Praxis in anderen Domänen, beispielsweise in der Medizin, bereits effektiv eingesetzt, stellen jedoch im Kontext professioneller Kompetenzen von Mathematiklehrkräften einen vergleichsweise neuen Ansatz dar. In die Simulation wurden unterschiedliche Formen von Scaffolding integriert, die sich als effektiv zur Förderung komplexer Kompetenzen erwiesen haben. Logdaten aus der Simulation dienten dazu, den Diagnoseprozess der angehenden Lehrkräfte zu analysieren. Diese Methode erlaubte es, den Diagnoseprozess in einer interaktiven Situation zu erfassen, die die Auswahl von Aufgaben und die Interpretation von Schülerlösungen beinhaltete.

Diese Dissertation umfasst zwei Publikationen, die zur Entwicklung wirksamer Fördermechanismen für die Diagnosekompetenzen angehender Lehrkräfte beitragen. Die erste Publikation evaluierte das Design und die Wirksamkeit einer digitalen Simulation mit unterschiedlichem Scaffolding zur Verbesserung der Diagnosegenauigkeit angehender Lehrkräfte. Die zweite Publikation analysierte Zusammenhänge zwischen den Dispositionen, dem Diagnoseprozess und der Diagnosegenauigkeit angehender Lehrkräfte, um die Implementierung adaptiven Scaffoldings zu unterstützen.

Die Ergebnisse zeigen, dass Simulationen sich für die Analyse und Förderung der Diagnosekompetenzen angehender Mathematiklehrkräfte eignen. Scaffolding, das auf das fachdidaktische Wissen fokussiert war, erwies sich als wirksamer für die Verbesserung der Diagnosegenauigkeit als Scaffolding, das auf Strategien im Diagnoseprozess fokussierte. Im Diagnoseprozess wurden systematische individuelle Unterschiede identifiziert, die für die Entwicklung adaptiven Scaffoldings genutzt werden können. Insgesamt kann dieses Forschungsprojekt als Beispiel für die Entwicklung und Evaluation von Simulationen zur Lehrkräfteprofessionalisierung während der universitären Ausbildung dienen.

Included and associated publications

Included publications

Schons, C., Obersteiner, A., Reinhold, F., Fischer, F., & Reiss, K. (2023). Developing a simulation to foster prospective mathematics teachers' diagnostic competencies: the effects of scaffolding. *Journal für Mathematik-Didaktik*, *44*(1), 59–82. https://doi.org/10.1007/s13138-022-00210-0

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2024). Toward adaptive support of pre-service teachers' assessment competencies: Log data in a digital simulation reveal engagement modes. *Learning and Instruction*, *94*, 101979. https://doi.org/10.1016/j.learninstruc.2024.101979

Associated publications

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2020). Förderung von Diagnosekompetenzen durch Scaffolding. In H.-S. Siller, W. Weigel, & J. F. Wörler (Eds.), *Beiträge zum Mathematikunterricht 2020* (p. 1535). WTM-Verlag. https://doi.org/https://doi.org/10.17877/DE290R-21544

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2022). Pre-service teachers' task selections for assessing students' misconceptions. 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* (p. 403, Vol. 4). PME.

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2023a). Teachers' diagnostic activities during task-based assessments in a digital simulation. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education* (pp. 171–178, Vol. 4). PME.

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2023b). Prozesse während der aufgabenbasierten Diagnose mathematischer Fehlvorstellungen in einer digitalen Simulation. In Goethe-Universität Frankfurt (Ed.), *Beiträge zum Mathematikunterricht 2022* (pp. 1197–1200). WTM-Verlag. https://doi.org/10.17877/DE290R-23407

Associated presentations

Schons, C. (2024a). Investigating mathematics teachers' assessment competencies: What can we learn from process data? *Atlantic Education Graduate Student Conference*. Fredericton (CA).

Schons, C., Strohmaier, A., Kadluba, A., & Obersteiner, A. (2024). Worauf achten Mathematiklehrkräfte bei der Diagnose von Schülerlösungen? Eine Analyse der Blickbewegungen. *11. Jahrestagung der Gesellschaft für Empirische Bildungsforschung (GEBF)*. Potsdam.

Schons, C. (2024b). Messung der Diagnosekompetenz angehender Mathematiklehrkräfte bei der Diagnose von systematischen Fehlern – Was können wir durch Prozessdaten lernen? *Vortrag in der Abteilung für Didaktik der Mathematik am IPN Kiel*. Kiel.

Schons, C., & Obersteiner, A. (2024). Adaptive prompting in simulation-based learning – COSIMA DiMa2. *Workshop Adaptivity in simulation-based learning*. Frauenchiemsee.

Schons, C. (2023). Analysing and fostering pre-service teachers' assessment competencies in a digital simulation. *Meeting of the Mathematics Pedagogy Group at the Center for Mathematical Cognition*. Loughborough (UK).

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2023). Diagnoseprozesse angehender Lehrkräfte in einer digitalen Simulation erklären Unterschiede in der Diagnoseakkuratheit. *10. Jahrestagung der Gesellschaft für Empirische Bildungsforschung (GEBF)*. Essen.

Schons, C., & Obersteiner, A. (2022). Aufgabenbasierte Diagnose mathematischer Fehlvorstellungen aus mathematikdidaktischer und psychologischer Perspektive. *Herbsttagung des Arbeitskreises Psychologie und Mathematikdidaktik der GDM*. Rauischholzhausen.

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2022a). Identifying students' misconceptions in a digital simulation. *12th International Conference of the EARLI SIG Conceptual Change*. Zwolle (NL).

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2022b). Task-based assessment of students' mathematical misconceptions in a digital simulation. *Analyzing and Integrating Process Data in Educational Research. Workshop des Scientific Network "Developing and stimulating competencies: Methodological challenges and opportunities for research"*. Loughborough (UK).

Schons, C., Obersteiner, A., Reinhold, F., Fischer, F., & Reiss, K. (2021b). Stimulating prospective mathematics teachers' diagnostic competencies with scaffolding. *Meeting of the scientific network "Developing and stimulating competencies: Methodological challenges and opportunities for research"*. Leuven (BE).

Schons, C., Obersteiner, A., Reinhold, F., Fischer, F., & Reiss, K. (2021a). Förderung von Diagnosekompetenzen angehender Mathematiklehrkräfte mit Simulationen: Effekte durch Scaffolding. *Sitzung des Arbeitskreises Empirische Bildungsforschung der GDM*. Online.

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2021a). Learning to diagnose: Effects of scaffolding in a simulation for prospective primary school teachers. *19th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI)*. Gothenburg (SWE).

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2021b). Scaffolding zur Förderung von Diagnosekompetenzen angehender Grundschullehrkräfte in Mathematik. *Cosima meets ProSim Symposium*. Online.

Contents

Acknowledgements ... i

Abstract ... ii

Zusammenfassung ... iii

Included and associated publications ... v

1 Introduction ... 1

2 Mathematics teachers' assessment competencies ... 5
- 2.1 Conceptualization ... 6
  - 2.1.1 Dispositions ... 7
  - 2.1.2 Assessment process ... 8
  - 2.1.3 Assessment performance ... 9
- 2.2 Assessment of student competencies and misconceptions in mathematics ... 9
- 2.3 State of research ... 12
  - 2.3.1 Task-related assessments ... 12
  - 2.3.2 Person-related assessments ... 13
- 2.4 Summary ... 14

3 Simulations for fostering mathematics teachers' professional competencies ... 16
- 3.1 Design of simulations ... 16
- 3.2 Scaffolding ... 17
- 3.3 Adaptivity ... 18
- 3.4 Summary ... 20

4 Log data in digital simulations ... 22
- 4.1 Benefits of log data to observe the assessment process ... 22
- 4.2 Interpretation of log data ... 23
- 4.3 Process indicators in assessments of student competencies and misconceptions ... 25
- 4.4 Summary ... 26

5 The present research ... 27

6 Summary of publications ... 29
- 6.1 Evaluation of design and effectiveness of a digital simulation with scaffolding (Paper A) ... 29
  - 6.1.1 Research goals ... 29
  - 6.1.2 Methodology ... 29
  - 6.1.3 Main findings ... 31

6.2 Identification of individual differences in the assessment process and associations with dispositions and assessment accuracy (Paper B) ... 32
6.2.1 Research goals ... 32
6.2.2 Methodology ... 32
6.2.3 Main findings ... 33
6.3 Associated publications ... 35

7 Discussion ... 37
7.1 Design and effectiveness of simulations for training mathematics teachers' assessment competencies ... 37
7.1.1 Design of the simulation ... 38
7.1.2 Effectiveness of the simulation with scaffolding ... 39
7.2 The relationship between dispositions and the assessment process in the genesis of assessments ... 40
7.2.1 Use of log data to analyze the assessment process ... 40
7.2.2 Development of adaptive scaffolding ... 41
7.3 Methodological considerations ... 42
7.4 Limitations and future directions ... 44

8 Conclusion ... 46

References ... 46

Appendix ... 54

1. Introduction

Student assessment is a central component of mathematics teachers' core practices (Bardy et al., 2024; Grossman, 2018; Kultusministerkonferenz, 2004). The quality of teachers' assessments affects students' learning outcomes in the classroom (Anders et al., 2010; Behrmann & Souvignier, 2013). In their daily classroom practice, teachers elicit and assess students' mathematical thinking by posing the right tasks and evaluating student solutions to make decisions on tailored follow-up instruction (Heitzmann et al., 2019). To meet these requirements, teachers need assessment competencies.

Assessment competencies are universally important in mathematics teaching. The following excerpt from Plato's *Meno* gives an example of a teacher-student interaction that highlights the need for assessment competencies in mathematics teaching (Plato and Jowett, 1999, pp. 32–35). Although our understanding of how teachers can effectively stimulate students' mathematical thinking has evolved over the decades, this excerpt underlines the importance of carefully selecting mathematical questions and the need for teachers to continuously react to students' responses. The scene depicts a situation in which Socrates (469–399 BC) is interacting with Meno (423–400 BC) and one of Meno's attendants. Socrates, as usual, takes on the role of the teacher, while Meno's attendant can be regarded as mathematics learner. Socrates poses the problem of finding the side length of a square that has twice the area of a given square (Figure 1):

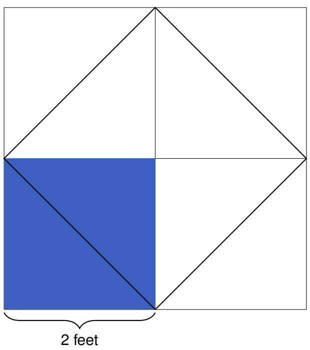

Figure 1 Illustration of the problem posed by Socrates. The side length of a square that has twice the area of the blue square shall be determined.

SOCRATES: And might there not be another square twice as large as this, and having like this the lines equal?

ATTENDANT: Yes.

SOCRATES: And of how many feet will that be?

ATTENDANT: Of eight feet.

SOCRATES: And now try and tell me the length of the line which forms the side of that double square: This is two feet – what will that be?

ATTENDANT: Clearly, Socrates, it will be double.

SOCRATES: Do you observe, Meno, that I am not teaching anything, but only ask him questions; and now he fancies that he knows how long a line is necessary in order to produce a figure of eight square feet; does he not?

MENO: Yes.

SOCRATES: And does he really know?

MENO: Certainly not.

SOCRATES: He only guesses that because the square is double, the line is double.

[...]

SOCRATES: Very good; I like to hear you say what you think. And now tell me, is not this a line of two feet and that of four?

ATTENDANT: Yes.

SOCRATES: Then the line which forms the side of eight feet ought to be more than this line of two feet, and less than the other of four feet?

ATTENDANT: It ought.

SOCRATES: Try and see if you can tell me how much it will be.

ATTENDANT: Three feet.

SOCRATES: Then if we add a half to this line of two, that will be the line of three. Here are two and there is one; and on the other side, here are two also and there is one: And that makes the figure of which you speak?

ATTENDANT: Yes.

SOCRATES: But if there are three feet this way and three feet that way, the whole space will be three times three feet? That is evident.

SOCRATES: And how much are three times three feet?

ATTENDANT: Nine.

SOCRATES: And how much is the double of four?

ATTENDANT: Eight.

SOCRATES: Then the figure of eight is not made out of a line of three?

ATTENDANT: No.

SOCRATES: But from what line? – Tell me exactly; and if you would rather not reckon, try and show me the line.

ATTENDANT: Indeed, Socrates, I do not know.

SOCRATES: Do you see, Meno, what advances he has made in his power of recollection? He did not know at first, and he does not know now, what is the side of a figure of eight feet: But then he thought that he knew, and answered confidently as if he knew, and had no difficulty; now he has a difficulty, and neither knows nor fancies that he knows.

[...]

SOCRATES: Mark now the farther development. I shall only ask him, and not teach him, and he shall share the enquiry with me: And do you watch and see if you find me telling or explaining anything to him, instead of eliciting his opinion. Tell me, is not this a square of four feet which I have drawn?

ATTENDANT: Yes.

[...]

SOCRATES: And this space is of how many feet?

ATTENDANT: Of eight feet.

SOCRATES: And from what line do you get this figure?

ATTENDANT: From this.

SOCRATES: That is, from the line which extends from corner to corner of the figure of four feet?

ATTENDANT: Yes.

SOCRATES: And that is the line which the learned call the diagonal. And if this is the proper name, then you are prepared to affirm that the double space is the square of the diagnonal?

ATTENDANT: Certainly, Socrates.

The scene highlights the importance of posing targeted questions and considering students' current mathematical competencies in teacher-student interactions. Activities such as posing problems and interpreting student solutions are core practices for mathematics teachers and part of their assessment competencies (Grossman, 2018).

Developing mathematics teachers' assessment competencies that enable them to master activities such as posing the right tasks is a challenge for mathematics teacher education (Heitzmann et al., 2019; Leuders et al., 2022; Sommerhoff et al., 2022). At the same time, there is a lack of practice opportunities for teachers that help them to systematically engage in these activities.

The aim of this dissertation project was to develop an evidence-based practice opportunity for mathematics teachers that effectively fosters their assessment competencies. To this end, we used a digital simulation of an assessment situation involving posing tasks and interpreting student solutions. We evaluated the design and effectiveness of the simulation for fostering mathematics teachers' assessment competencies by focusing specifically on the assessment process. In digital simulations, the assessment process can be observed in a unique way through the log data. By investigating how to effectively foster teachers' assessment competencies with digital simulations, this research should contribute to the evidence-based development of teachers' professional competencies.

2. Mathematics teachers' assessment competencies

The importance of teachers' competencies in assessing student learning has been acknowledged by policy makers in teacher education. In Germany, for example, common standards for teachers' professional competencies have been established (Kultusministerkonferenz, 2004). These standards include the broad construct of *assessment competencies*, which comprises the competencies of posing appropriate questions and tasks and of interpreting student solutions when interacting with individual students.

Similar to the established standards in Germany, the competencies of *eliciting and interpreting student thinking* and *diagnosing particular common patterns of student thinking and development in a subject-matter domain* have been established in the internationally beknown core practices in teacher education (Grossman, 2018). Taken together, developing competencies of posing appropriate questions and interpreting students solutions to these question to make tailored instructional interventions is a central requirement in international mathematics teacher education. It enables teachers to assess individual students' mathematical thinking and to provide student-centered, tailored instruction.

Empirical studies support the importance of teachers' assessment competencies for student learning: The accuracy of mathematics teachers' assessments was positively related to student performance (Anders et al., 2010; Behrmann & Souvignier, 2013). At the same time, teachers' often struggle to make accurate assessments, with a correlation of $r = .63$ between teachers' assessments of student performance and their actual performance (Südkamp et al., 2012). These results point at a potential for improving teachers' assessment accuracy, and thereby improving student learning outcomes in the classroom.

Despite the importance of teachers' assessment competencies, it remains largely unclear how mathematics teachers' development of assessment competencies can be effectively fostered (Südkamp & Praetorius, 2017). Research on this issue was often not specific to domains, providing insights into the relation between assessment accuracy and student learning outcomes on a general level. However, evidence on how to support the development of competencies needed in specific situations that occur in mathematics instruction is lacking. For example, selecting mathematical tasks to assess students' mathematical competencies and misconceptions requires specific knowledge about task potentials and typical student errors in specific tasks. Moreover, studies have often analyzed the accuracy of teachers' assessment in isolation instead of including teachers' dispositions and their processes that lead to their assessment (Kolovou et al., 2024). Taken together, a domain-specific perspective, including individual teachers' dispositions and assessment process seem necessary for developing effective support of mathematics teachers' assessment competencies.

This chapter provides a conceptualization of mathematics teachers' assessment competencies that is based upon Blömeke et al.'s (2015) comprehensive framework of professional competencies. It goes beyond the accuracy of assessments and includes teacher dispositions as well as the assessment process. In this dissertation project, this conceptualization serves as the basis for analyzing and fostering mathematics teachers' assessment competencies. We will then focus on the assessment of individual students' competencies and misconceptions in mathematics, which is particularly important for providing student-centered instruction and tailored support for students. Finally, this chapter gives a brief review of research on mathematics teachers' assessment competencies.

2.1 Conceptualization

Teacher assessments can be categorized into three different types (Urhahne & Wijnia, 2021): *person-related assessments* that aim to generate knowledge about individual students' competencies in a domain, *task-related assessments* that aim to generate knowledge about task characteristics, and *person-specific assessments* that aim to generate knowledge about students' personal characteristics, such as giftedness or hyperactivity disorder. This project focuses on person-related assessments in mathematics, particularly on assessments of students' mathematical competencies and misconceptions. Accordingly, we define assessment competencies as a set of skills for accurately assessing students' learning processes and outcomes to initiate effective instruction (Artelt & Gräsel, 2009).

Models of teachers' assessment competencies include individual dispositions, such as professional knowledge, and performance in practice situations (Heitzmann et al., 2019; Herppich et al., 2018). Blömeke et al.'s (2015) competence model additionally includes so-called situation-specific skills to bridge the gap between dispositions and performance in practice situations. In this project, we apply this comprehensive view of competencies to assessment situations. This enables us to not only describe teachers' accuracy or knowledge, but to understand the complex mechanisms of how dispositions translate to accurate assessments in authentic situations. In the following, the components of the model are explained in greater detail.

2.1.1 Dispositions

Mathematics teachers' *dispositions* are commonly subdivided into cognitive and affective-motivational dispositions (Heitzmann et al., 2019; Figure 2). Cognitive dispositions include different facets of professional knowledge: content knowledge (CK, knowledge of the subject matter), pedagogical knowledge (PCK, knowledge about teaching) and pedagogical content knowledge (PK, knowledge about how to teach specific subject matter) (Depaepe et al., 2013; Hill et al., 2005; Shulman, 1986). In mathematics-specific assessment situations (such as assessing mathematical competencies and misconceptions), content knowledge and pedagogical content knowledge are particularly important, whereas general pedagogical knowledge unrelated to the content plays only a minor role.

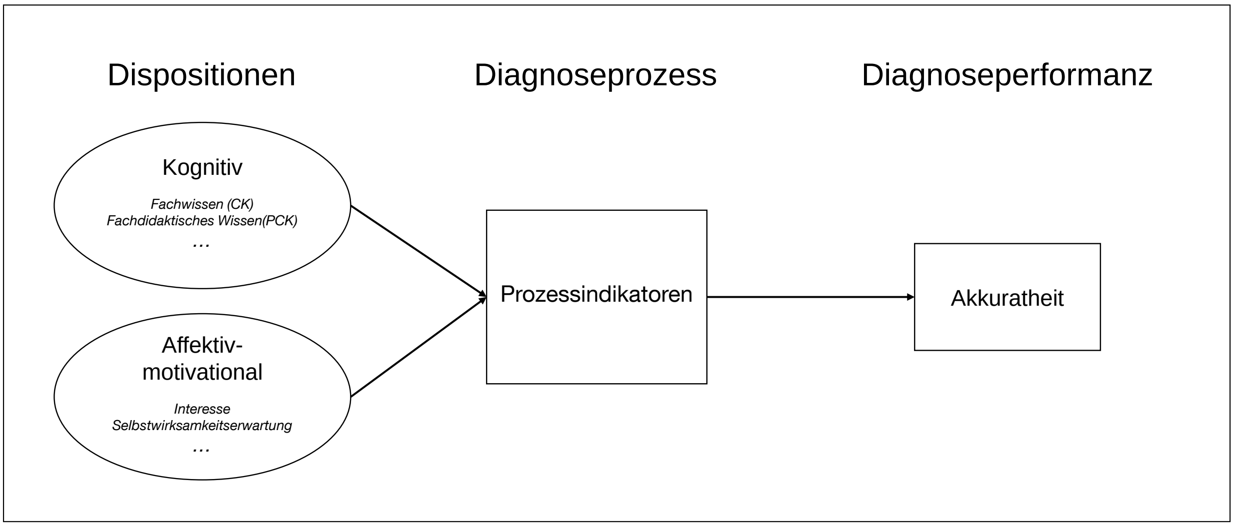

Figure 2 Competency framework based on Blömeke et al. (2015) applied to assessment situations.

Affective-motivational dispositions are more diverse and there is less consensus about which affective-motivational dispositions are relevant in assessment situations. In research on professional competencies across domains, there is consensus that individual interest and self-efficacy are important dispositions (Herppich et al., 2018). Individual interest is commonly defined as a person's relatively stable orientation of an individual toward an object or activity (Schiefele, 1991). Self-efficacy refers to persons' beliefs about their ability to master a task (Bandura, 1977). These dispositions are crucial for mastering assessment situations, and it remains unclear how they transfer to the assessment process and the accuracy of assessments.

2.1.2 Assessment process

The process that takes place while a teacher is generating an assessment is referred to as *the assessment process*. Depending of the assessment situation, various models are used to derive indicators that determine the quality of the assessment process. What these models have in common is that they focus on the assessment process, offering a theoretical foundation for explaining variance in teachers' assessment accuracy.

Loibl et al. (2020) developed a cognitive process model that is primarily used to investigate the assessment process in task-related assessments. The assessment process is conceptualized as cognitive information processing, consisting of the processes *perceiving*, *interpreting*, and *decision-making*. In research using this model, indicators of the quality of the assessment process included, for example, fixation durations on specific task features or reaction times in assessment items (Brunner et al., 2024; Rieu et al., 2022).

When videos are used to support assessment competencies, the professional vision model is often the basis for analyzing the assessment process (Sherin, 2001). This model is based on the assumption that observing professional situations (e.g., assessment situations) can support the acquisition of competencies. The professional vision model distinguishes between *noticing* and *knowledge-based reasoning* (Seidel & Stürmer, 2014). Accordingly, indicators of the assessment process are noticing of relevant cues and describing and explaining them appropriately (Codreanu et al., 2021).

The process model by Heitzmann et al. (2019) focuses on activities that occur during the assessment process in person-related assessments that involve teacher-student interaction. The model is based on Fischer et al.'s (2014) framework for scientific reasoning and outlines specific *diagnostic activities* involved in the process of generating new knowledge (e.g., about student's mathematical competencies). Activities during the assessment process include *generating evidence*, *evaluating evidence*, *stating hypotheses*, and *drawing conclusions*. Since we investigated a person-related assessment situation that involved teacher-student interaction, Heitzmann et al.'s (2019) model was used to operationalize the assessment process. Applied to situations in which teachers pose tasks and interpret student solutions to assess students' mathematical competencies, the activities in this model can be specified: *generating evidence* refers to posing tasks, and the activities *evaluating evidence*, *stating hypotheses* and *drawing conclusions* refer to teachers' interpretations of student solutions. *Evaluating evidence* involves describing observable features within individual task solutions, which may indicate mathematical competencies or misconceptions. *Stating hypotheses* refers to statements about a student's potential competencies or misconceptions that contain expressions of uncertainty. *Drawing conclusions* goes beyond simply evaluating evidence and involves synthesizing information from multiple task solutions to make general statements about a student's competencies or misconceptions.

2.1.3 Assessment performance

Assessment performance refers to the quality of teachers' generated assessment and is indicated by the accuracy of assessments (Hoge & Coladarci, 1989; Südkamp et al., 2012; Urhahne & Wijnia, 2021). Assessment accuracy is defined as the degree to which an assessment corresponds to a predefined, objective norm. This norm can be based on students' actual performance on tests that teachers need to predict (Südkamp et al., 2012), or it can be a more comprehensive norm, such as students' competence levels derived from a competence model (Reiss & Obersteiner, 2019).

Assessment accuracy is crucial, as it serves as the basis for instructional decisions (Heitzmann et al., 2019) and is related to student performance (Anders et al., 2010; Behrmann & Souvignier, 2013). Therefore, the ultimate goal of fostering teachers' assessment competencies is to increase the accuracy of their assessments.

2.2 Assessment of student competencies and misconceptions in mathematics

Teachers' accurate assessment of students' mathematical competencies and misconceptions is the basis for making instructional decisions and is therefore crucial for student-centered and adaptive instruction (Corno, 2008; Hardy et al., 2019). For several mathematical topics, such as primary school arithmetic, students' competencies are described in terms of different competence levels in comprehensive competence models (Reiss & Winkelmann, 2009). Although competence models usually describe student competencies on a rather general level, teachers can use such models as a reference when making assessments (Reiss & Obersteiner, 2019). Competence models establish a benchmark against which teachers can evaluate student outcomes, enabling a more objective assessment of student competencies compared to relative comparisons with class average, for instance. Thus, they can assist teachers during the assessment process and contribute to more accurate assessments. Given that objective norms were often not present in previous studies and that teachers tend to make more accurate assessments when a reference is available, competence models can assist teachers in the assessment process and enable them to make more accurate assessments (Südkamp et al., 2012).

Table 1

Summary of the competency model by Reiss and Winkelmann (2009) for the topic numbers and operations.

Competence level	Description of competencies
1	Understanding the structure of the decimal system; addition and multiplication of numbers from 1 to 10; doubling of larger numbers
2	Understanding the structure of the decimal system in various representations; identifying simple patterns; using basic arithmetic to solve simple word problems
3	Proficient writing and reading of numbers in various representations; correct usage of the number 0; expanded addition and subtraction of larger numbers; understanding the relation between addition and subtraction as well as between multiplication and division; modelling and solving simple real world situations
4	Mastering all basic arithmetic operations; using relations between different operations, modelling and solving more complex real world situations
5	Solving complex tasks with different strategies; solving difficult equations; calculating with decimals

An effective way to assess students' mathematical competencies in practice is to pose questions or tasks (Black & Wiliam, 2009). Given a reference, such as a competence model (Table 1), tasks can be assigned to different levels that describe specific competencies, such as *understanding the structure of the decimal system*. A student solution to a task that focuses on this competency can reveal evidence that facilitates instructional decisions. For example, understanding the structure of the decimal system forms a very basic competency (Table 1). Consequently, teachers might begin an assessment with tasks asking a student to use different representations of decimal numbers, such as place value representations or symbolic representations. If a student does not master these tasks, it might be effective to provide more instruction on the decimal system before moving to more advanced topics. Conversely, if a student has a good understanding of the decimal system, the teacher could use more advanced tasks related to competencies like *mastering all basic arithmetic operations* to assess the students' current competency level.

To assess students' current knowledge of a mathematical concept, it is important for teachers to be aware not only of students' competencies, but also of their misconceptions (Larrain & Kaiser, 2022; Vosniadou & Verschaffel, 2004). For fundamental mathematical topics such as basic arithmetic or fractions, misconceptions are well described in the literature (Eichelmann et al., 2012; Padberg & Benz, 2021; Radatz, 1980a, 1980b). Misconceptions are thought to be robust over time, as reflected in consistent patterns of errors on specific tasks (Vosniadou & Verschaffel, 2004). Based on a given misconception, one can decide for each task whether it has the potential to reveal a typical error related to the misconception.

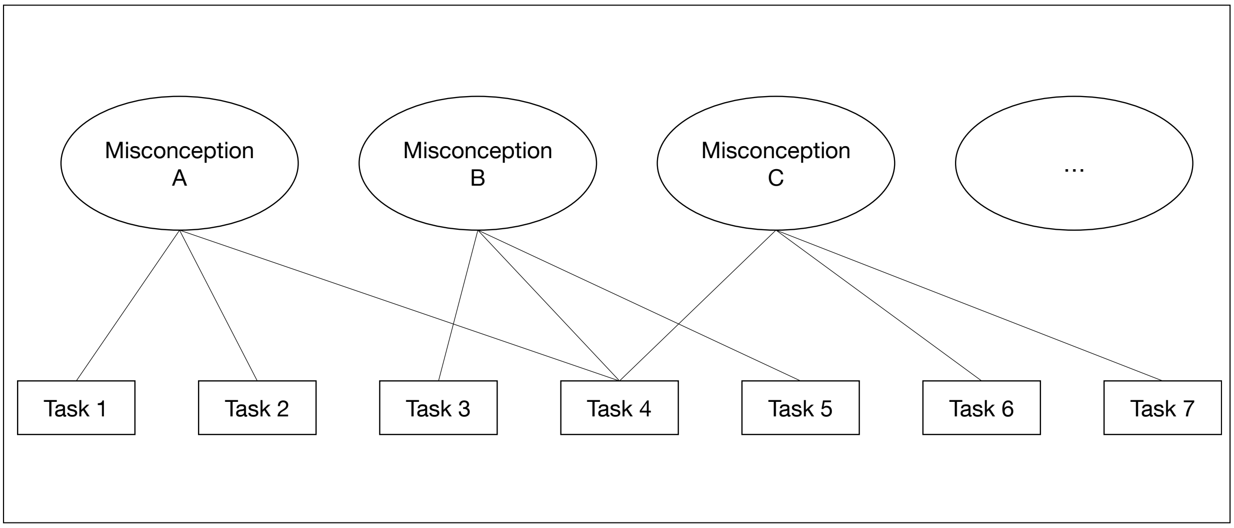

Figure 3 Illustration of how tasks can be used to elicit misconceptions. A connection between a task and a misconception means that the task has the diagnostic potential to reveal an error related to the misconception. A student with one of these misconceptions would make a consistent pattern of errors in the connected tasks.

Figure 3 illustrates in a simplified way how students' misconceptions can be assessed based on errors in task solutions. If the tasks are classified in terms of their *diagnostic* *potential* to reveal errors related to a misconceptions, consistent patterns of errors in these tasks allow to infer an underlying misconception.

For making accurate assessments of mathematical competencies and misconceptions, teachers' need to be able to pose the right tasks and to interpret student solutions appropriately (Heitzmann et al., 2019). Mastering the processes of selecting tasks and interpreting student solutions determines the quality of the assessment process in this specific assessment situation, which in turn may determine the accuracy of the final assessment.

2.3 State of research

Research on mathematics teachers' assessment competencies focused on task-related assessments and person-related assessments (Heitzmann et al., 2019; Leuders et al., 2022; Loibl et al., 2020). Research on task-related assessments has mainly focused on cognitive processes when assessing task characteristics, most prominently on task difficulty. The main research goal of this line of research is to explain the cognitive mechanisms during the genesis of task-related assessments. To observe indicators of cognitive processes, commonly used methods in experimental studies include eye tracking and think aloud.

Research on person-related assessments in authentic practice situations focused on more complex assessment situations involving teacher-student interaction. The main research goal of this line of research is to describe and explain teacher actions in practice situations to facilitate the acquisition of assessment competencies. Methods include classroom observations, videos and simulations of practice situations (Leuders et al., 2022).

2.3.1 Task-related assessments

Studies in mathematics education that have focused on task-related assessments primarily investigated pre-service teachers' perception and interpretation of task difficulty (Brunner et al., 2024; Ostermann et al., 2018; Rieu et al., 2022). In experimental studies, the effects of specific PCK interventions on pre-service mathematics teachers' assessment performance were analyzed. Across different mathematical content domains, PCK interventions led to higher assessment performance in perceiving and interpreting task difficulties. Indicators of the assessment process were response times (Rieu et al., 2022) and eye movements

(Brunner et al., 2024). Main findings were that the fixation durations on relevant task features were related to the accuracy. For the PCK intervention, it was found that a PCK intervention lead to longer fixation durations on relevant task features and to more accurate interpretations (Brunner et al., 2024). In addition, a PCK intervention was causally related to assessment accuracy in identifying and interpreting task features (Rieu et al., 2022).

The reported studies used Loibl et al.'s (2020) model (Section 2.1.2) to derive indicators of the assessment process. The use of eye movement parameters as indicators of the assessment process had the advantage of revealing qualitative differences related to specific cognitive processes. For instance, it allows for distinguishing whether pre-service teachers focus on relevant features of the tasks or on aspects that are irrelevant for the assessment.

Taken together, the studies provide evidence that pre-service teachers' PCK is causally related to their assessment process and their accuracy in task-related assessments. Limitations of the studies were that they did not account for teachers' affective-motivational dispositions and focused only on the knowledge facet PCK instead of considering other knowledge facets such as CK.

2.3.2 Person-related assessments

Studies on person-related assessments in mathematics used different approximations of practice to provide authentic assessment situations (Grossman et al., 2009). For example, video-vignettes were used to study teachers' assessment process within the professional vision framework (Section 2.1.2). In particular, pre-service teachers' noticing of relevant cues and their reasoning were analyzed (Codreanu et al., 2020; Nickl, Sommerhoff, Böheim, et al., 2024). Results indicated that pre-service teachers' reasoning was mostly superficial, describing overt observations. Wildgans-Lang et al. (2020) and Larrain and Kaiser (2022) reported similar findings that indicated that pre-service teachers struggled to state hypotheses and instead described only superficial features in student solutions. This suggests that pre-service teachers need support to improve their reasoning during the assessment process. A limitation in the reported studies was that they did not investigate whether pre-service teachers' reasoning quality during the assessment process was related to their dispositions and their assessment accuracy.

In assessments of individual students competencies in role play simulations, a positive correlation between CK and assessment accuracy was found (Kron et al., 2021, 2022). CK was also related to the quality of the assessment process, operationalized as the relative frequency of selected high quality tasks during the interaction with the students. Kron et al. (2022) also found that the interaction of individual interest in mathematics education and PCK was related to the quality of task selection. This effect was interpreted as suggesting that interest serves as a prerequisite for knowledge activation. The results indicate that affective-motivational variables play an important role in the assessment process in authentic practice situations. In these studies, the operationalization of the assessment process focused primarily on task selection and found pronounced individual differences in pre-service teachers quality of task selections (Kron et al., 2021). Focusing solely on the process of task selection without considering teachers' interpretations of student solutions is a limitation as only part of the teacher-student interaction is captured. Furthermore, interpreting the influence of interest could rely on more established theories and constructs.

2.4 Summary

This chapter introduced a comprehensive model for investigating teachers' assessment competencies. The model includes cognitive and affective-motivational dispositions, the assessment process, and assessment performance. Thus, it can facilitate our understanding of the mechanisms underlying accurate assessments, going beyond research that has mainly described the accuracy of teachers' assessments.

The state of research of mathematics teachers' assessment competencies is described along two lines of research: task-related assessments with a focus on cognitive processes, and person-related assessments in authentic practice situations. A key finding was that professional knowledge was related to the quality of the assessment process and assessment performance.

At least three research gaps can be identified: First, it remains unclear whether teachers' PCK is also related to the assessment process and accuracy in person-related assessments as it was found in task-related assessments. The role of teachers' CK remains unclear and was not considered in studies on task-related assessments. Second, the assessment process in interactive assessment situations has mostly been operationalized as the quality of task selection without holistically considering also the interpretation of student solutions to selected tasks. However, the quality of the interpretation of student solutions is the basis for the subsequent task selection and for instructional decisions. Third, it is yet an open question how effective instructional support of teachers' assessment competencies can be designed. Instructional support can focus on dispositions such as PCK, or it can focus on the assessment process itself, depending on how dispositions and the assessment process contribute to the genesis of accurate assessments.

To address these research gaps, we need authentic practice environments that allow for a comprehensive study of assessment competencies in a controlled setting. Simulations offer a promising approach and create a controlled research setting while simultaneously providing an authentic practice situation (Fischer et al., 2022; Grossman et al., 2009). The next chapter explores the potential of simulations for analyzing and fostering teachers' professional competencies, including assessment competencies.

3. Simulations for fostering mathematics teachers' professional competencies

In many domains, such as medicine, simulations are used to prepare learners for real professional situations (Fanning & Gaba, 2007). Simulations are commonly defined as a system representing a segment of reality with possibilities for learners to interact with the system (Heitzmann et al., 2019). Simulations have at least three advantages compared to real situations in preparing learners for practice (Fischer et al., 2022): First, the tasks can be repeated to systematically monitor and enhance performance. Second, the situation can be simplified by only focusing on key aspects of the situation, omitting irrelevant aspects, and third, the task can be designed according to learners' needs, whereas in real situations the situations intended for learning may not occur at all.

Among educational disciplines, medical education in particular has a tradition in using simulations (Cook et al., 2012). In teacher training, simulations are less commonly used – despite their benefits and potential (Chernikova, Heitzmann, Fink, et al., 2020). To ensure that a simulation supports pre-service mathematics teachers' acquisition of core practices, it is important to consider instructional design principles. In this section, we will first describe instructional design principles for effective simulations and then focus on the potential of simulations to support the acquisition of professional competencies.

3.1 Design of simulations

The design of simulations can contribute to learners' experience and learners' performance while engaging in the simulation (Brom et al., 2017; Chernikova, Heitzmann, Stadler, et al., 2020). A high-quality learning experience requires learners' presence in the situation. Presence refers to the extent to which learners feel involved in the situation during the simulation. The concept of authenticity is similarly important for learners' experience and refers to the extent to which learners perceive a situation as similar to a real situation (Levin et al., 2023; Stürmer et al., 2024). Both presence and authenticity are relevant ensuring the ecological validity of a simulation, and ecological validity is a prerequisite for ensuring that the core practices learned in simulations are transferable to practice situations.

The quality of learners' experience also depends on the interactivity of the simulation (Cook et al., 2013). Simulations that provide learners with a greater degree of freedom to influence the course of the situation are more immersive (Dawley & Dede, 2014). Learners' performance is influenced by the extraneous cognitive load that learners are exposed to while engaging in the simulation (Renkl & Atkinson, 2003). Extraneous cognitive load refers to mental resources that are devoted to task-irrelevant features (Sweller, 2010). The availability of relevant information and the complexity of navigating through the situation can increase learners' extraneous load in simulations. To minimize learners' extraneous load, the design of simulations should contain as few task-irrelevant features as possible. However, this may also reduce the ecological validity and there is a trade-off between ecological validity and using task-irrelevant features when designing a simulation (Codreanu et al., 2020). Therefore, researchers need to find a balance between the ecological validity of the situation provided by the simulation and the extraneous cognitive load that learners are exposed to. Taken together, the design of a simulation influences learners' experience and performance to a large extent and the goal is to ensure a high-quality learner experience that supports high performance. Simulations allow for implementing additional support to enhance performance, and one support measure that has proven to be effective across learning domains is scaffolding.

3.2 Scaffolding

Scaffolding refers to a specific kind of instructional support that is applicable in complex and interactive situations (Wood et al., 1976). Scaffolding aims to support learners in solving complex tasks by reducing and regulating the complexity. The support is meant to intervene between the current level of knowledge and the level of knowledge to be acquired and aims to empower learners to master situations that they could not master based on their current knowledge level alone (Vygotskiĭ & Cole, 1978; Wright, 2018).

The effectiveness of scaffolding depends on learners' experience, type of scaffolding, and the methodology (Belland et al., 2017; Chernikova, Heitzmann, Stadler, et al., 2020): For inexperienced learners, effective scaffolding provides a high level of guidance and should be strongly related to the content of the learning task (Chernikova, Heitzmann, Fink, et al., 2020). In this case, scaffolding is most effective when it is presented continuously during the learning process. In contrast, for experienced learners, effective scaffolding provides less guidance, for example, by providing strategic rather than content-related support (de Jong & van Joolingen, 1998). In this case, scaffolding does not need to be presented permanently but can also fade out during the learning process (Belland et al., 2017).

Since these principles have primarily been derived from studies in complex problem-solving and medical education, it remains unclear whether the same principles of effective scaffolding apply to simulations in teacher education. Furthermore, very few studies have compared different types of scaffolding within the same learning domain, making it unclear which type of scaffolding is most effective for developing teacher competencies, such as assessment competencies.

For teachers assessment competencies, scaffolding that enhances teachers' assessment accuracy can focus on cognitive and affective-motivational dispositions, or on the assessment process (Figure 2). When focusing on cognitive dispositions, providing specific content knowledge has proven to be effective scaffolding (Brunner et al., 2024; Sommerhoff et al., 2023). However, for affective-motivational dispositions, it is less clear what effective scaffolding can focus on. In the assessment process, proficient learners are known to generate and verify hypotheses (Section 2.1.2), leading to well-founded conclusions (de Corte et al., 1991; de Jong & van Joolingen, 1998). At the same time, there is evidence that pre-service teachers struggle to generate hypotheses, and rather describe overt cues in diagnostic situations (Bauer et al., 2020; Codreanu et al., 2021; Larrain & Kaiser, 2022; Wildgans-Lang et al., 2020). Therefore, effective scaffolding focusing on the assessment process could prompt learners to state hypotheses or draw conclusions.

These types of scaffolding can be implemented using a static or adaptive scaffolding methodology. While static scaffolding means that all learners receive the same predefined instructional support throughout the learning situation, adaptive scaffolding can take into account learners' individual needs and can change during the learning situation.

3.3 Adaptivity

In the context of digital learning environments, adaptivity refers to the ability of a learning environment to evaluate relevant learner variables to adjust to individual learners with the goal of enhancing learning outcomes (Plass & Pawar, 2020). By adjusting, the learning environment provides individual learners with exactly the learning experience they need.

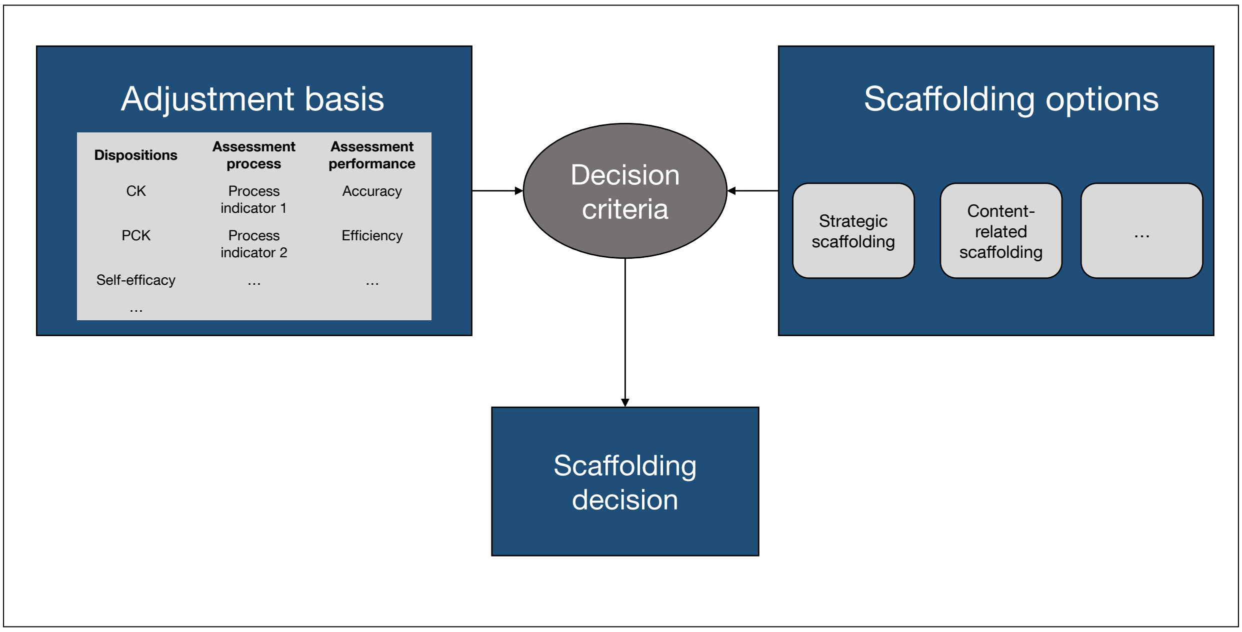

Figure 4 Overview of adaptive scaffolding: The adjustment basis is used to identify relevant individual differences among learners. Based on these differences, targeted scaffolding is selected and provided from a set of available scaffolding options.

The set of learner variables that the system evaluates forms the adjustment basis. For the adjustment basis, specific decision criteria are defined, according to which a particular type of scaffolding is presented to individual learners in a given situation from a set of scaffolding options (Figure 4). The adjustment can occur with varying temporal granularity (Tetzlaff et al., 2021): A learning environment that adapts at the macro-level evaluates learner variables prior to the learning task and uses this assessment as a basis for adjustment to make decisions for presenting scaffolding for the learning task. In this case, the adjustment is determined *before* the start of the learning task and is not further modified during the learning task. A learning environment that adapts on the micro-level evaluates learner variables in real time *during* the learning task, and the learner variables are updated and used as the adjustment basis during the learning task.

In simulations for fostering teachers' assessment competencies, the adjustment basis can include dispositions, assessment process, or assessment performance. Since dispositions are assumed to be stable over time, it is reasonable to measure them before the assessment situation and use them for macro-adaptive support. Indicators during the assessment process can only be measured while teachers are engaged in an assessment situation and are likely to change over time. Consequently, these indicators should be measured frequently and could be used for micro-adaptive scaffolding, that is, scaffolding that is adapted in real time based on the measured indicators. Assessment performance could be used for adaptation to track the learning progress of the teachers, to provide scaffolding in the form of the most appropriate follow up assessment situations (Belland et al., 2017).

Simulations with adaptive scaffolding are more effective than simulations with static scaffolding since they provide the most effective learning experience for each learner at each time. However, creating decision rules to identify which type of scaffolding is most effective for which learners poses a significant challenge for research.

To derive decision criteria to support mathematics teachers' assessment competencies adaptively, it is important to understand how teachers' dispositions and indicators of the assessment process contribute to accurate assessments. To this end, Plass and Pawar (2020) outlined key steps to help derive decision criteria, that can then be used to implement adaptive scaffolding in the next step. The first and most basic step is to ensure reliable measurement of the learner variables involved. While there are well-established measurement instruments for assessing pre-service teachers' dispositions and assessment performance, reliably and holistically measuring the assessment process itself remains challenging. Second, variables need to be identified that should be included in the adjustment basis. Finally, we need evidence about what type of scaffolding leads to higher assessment performance based on individual learner variables. These steps provide the knowledge needed to generate decision criteria and scaffolding that can most effectively foster pre-service teachers' assessment performance.

3.4 Summary

In this chapter, we outlined the benefits of simulations as a tool for analyzing and fostering mathematics teachers' professional competencies. We discussed design principles that make simulations both ecologically valid and effective for learning. Different kinds of scaffolding can be implemented as instructional support in digital simulations to further enhance teachers' assessment performance, but it is still unclear which types of scaffolding are effective for fostering pre-service teachers' assessment competencies. The implementation of scaffolding can be adaptive by providing different types of scaffolding based on individual learner variables. It remains a challenge to derive specific decision criteria to determine which learner should receive what kind of scaffolding in a specific situation. To make progress in this regard, a first step is to measure the assessment process holistically. For this purpose, log data from digital simulations can be used.

4. Log data in digital simulations

With the increasing use of digital technology for learning, log data have gained importance in educational research (Goldhammer et al., 2021). The term log data refers to user data in digital environments that are automatically recorded as users engage with the system (Greiff et al., 2016). A major advantage of log data is that it can capture not only learning outcomes, but also the entire learning process. The potential of log data has been shown in studies across educational domains and contexts (Ferguson, 2012; Fouh et al., 2014; Reinhold et al., 2020; Theobald et al., 2018). Accordingly, log data also promise insights for research on teacher competencies with digital simulations.

This chapter discusses the potential of log data for studying mathematics teachers' assessment processes with digital simulations. First, we focus on the general benefits of using log data to observe learning processes. Then, the chapter addresses challenges in interpreting log data and finally, we outline process indicators that can be derived from log data in assessments of individual students.

4.1 Benefits of log data to observe the assessment process

There are several advantages to using log data in research on learning and instruction. For the study of mathematics teachers' assessment competencies with digital simulations, this section highlights three key benefits of log data.

First, log data can provide insights into the assessment process, not just the accuracy of the assessment. This is particularly important for understanding the mechanisms that lead to accurate assessments – one of the main research goals in this research area and in this project (Chapter 2).

Second, log data are recorded unobtrusively and in real time. Learners' progression remains uninterrupted, thereby increasing external validity. The timing of activities within the assessment process can be accurately captured through timestamps in the log data, making log data superior to other methods for measuring the duration of activities in the learning process (Kovanovic et al., 2016).

Third, log data can be used not only to observe cognitive processes in highly controlled digital environments, but also to observe more complex behavioral processes in interactive digital environments such as simulations (Azevedo et al., 2013; Greene & Azevedo, 2010). A key characteristic of simulations is that the process can be navigated by the learner rather than being predetermined (Section 3.1). As authentic assessment situations are by nature more complex and interactive, using log data from digital simulations is advantageous.

Some of these benefits also apply to other methods that are commonly used in educational research to assess learning processes, for example think aloud or eye-tracking. However, think aloud requires participants to use cognitive resources to observe their own process, which may affect their engagement in the task and consequently the external validity (Cotton & Gresty, 2006). This is different with eye-tracking, but eye-tracking is best suited for observing cognitive processes in very controlled laboratory settings rather than processes in interactive and more complex assessment situations (Strohmaier et al., 2020). Taken together, log data provide a unique way to investigate mathematics teachers' assessment processes in digital simulations.

4.2 Interpretation of log data

Digital learning environments are used in a variety of educational disciplines with different research goals (Haleem et al., 2022; Wang, 2021). As a result, there is a great variety in the types of data that these learning environments assess through the log data (Goldhammer et al., 2021; Haleem et al., 2022). Since there is no universal way to use these different types of log data, the interpretation poses challenges for researchers. Researchers need to interpret log data meaningfully within their discipline and context (Goldhammer et al., 2021). This section presents selected examples to illustrate the use of log data from digital learning environments for educational research.

A group of studies used rather generic process indicators based on log data to analyze solution processes, most prominently the time learners spent while solving a task (time on task). Although relationships between general process indicators such as time on task and the learning outcome were sometimes found (Goldhammer et al., 2014; Scherer et al., 2015), it is challenging to draw conclusions that contribute to educational theories. For example, learners' time on task can have entirely different meanings: A short reaction time can indicate that the learner is working through a task very efficiently, or it could mean that the learner lacks the knowledge to solve the task and is simply skipping the task. A long reaction time may indicate that the task was solved efficiently but the result was double-checked, or it may suggest that the solution process itself was inefficient. The ambiguity of general measures such as time on task is particularly evident in less controlled environments, such as in digital simulations of interactive learning situations.

Another group of studies used log data to derive more specific process indicators. For example, log data were used to infer strategies in complex problem-solving tasks in a digital assessment (Greiff et al., 2016; Stadler et al., 2020). Log data enabled to distinguish whether learners applied the efficient vary-one-thing-at-a-time strategy (VOTAT) or a less efficient trial-and-error strategy during the solution process. Students using VOTAT apply the principle of isolating one variable at a time when exploring a problem. This approach allows them to identify the effect of each individual element within the problem. In this example, log data allowed a systematic investigation of the relationship between using VOTAT during the solution process and problem-solving performance. This shows how log data can be used to generate unique insights that would be difficult to observe in the same quality with other methods.

In mathematics education, Reinhold et al. (2020) used log data from writing-to-learn tasks in a digital textbook on fractions. Here, the combination of multiple process indicators (e.g., number of words) was interpreted as cognitive and behavioral engagement. Differences in the quality of students' engagement could be identified based on patterns of process indicators, and one main finding was that the process patterns derived from log data predicted the learning outcome above and beyond learners' knowledge. This study showed the potential of log data for research in mathematics education and also illustrated how a more complex solution process can be captured and interpreted by using patterns in multiple process indicators.

In complex processes such as the assessment process involving teacher-student interaction, where multiple indicators are relevant, theory-driven interpretation is challenging. As demonstrated in the study by Reinhold et al. (2020), the cognitive engagement framework is suitable for interpreting patterns across multiple process indicators (Fredricks et al., 2004). The cognitive engagement framework also allows for hypotheses about the influence of affective-motivational dispositions on learning processes. For example. studies across various learning domains found a link between interest and engagement (Ainley et al., 2002). The ICAP model builds on the theory of cognitive engagement and further distinguishes qualitative differences in engagement modes (Chi & Wylie, 2014). For individual learning, it differentiates between *passive*, *active*, and *constructive* cognitive engagement. However, to use this model for a meaningful interpretation of log data, the process indicators that characterize the different engagement modes in the respective context must be specified. The following section specifies process indicators that can be applied in interactive, person-related assessments of students mathematical competencies and misconceptions.

4.3 Process indicators in assessments of student competencies and misconceptions

In assessments of student competencies and misconceptions based on tasks, the assessment process consists of the process of selecting assessment tasks and the process of interpreting student solutions to selected tasks (Section 2.2).

During the process of selecting assessment tasks, the number of tasks with and without diagnostic potential (Section 2.2) serves as a key indicator of the quality of the assessment process (Kron et al., 2021). Teachers with high quality task selection are able to select tasks strategically, which is reflected in a high proportion of tasks with diagnostic potential. Conversely, in a low quality task selection process, the diagnostic potential of tasks is not taken into account.

During the process of interpreting student solutions, teachers' reasoning activities serve as key indicators of quality (Codreanu et al., 2020; Fischer et al., 2014; Heitzmann et al., 2019). These include evaluating evidence from individual task solutions, generating hypotheses about students' competencies and misconceptions, and drawing conclusions (Section 2.1.2). Teachers with high quality task interpretation are able to state specific hypotheses and draw appropriate conclusions about students' mathematical competencies based on the generated evidence.

These indicators can capture the quality of the assessment process in an interactive assessment situation that involves the processes of selecting tasks and interpreting student solutions. A challenge is to develop a simulation that elicits teachers' processes of task selection and interpreting student solutions.

4.4 Summary

This chapter discussed the potential of log data for investigating mathematics teachers' assessment processes with digital simulations, highlighting several key advantages over other methods of assessing learning processes. Log data can provide insights beyond mere accuracy in interactive assessment situations in an unobtrusive way, enabling a deeper understanding of the mechanisms underlying accurate assessments.

Interpreting log data, however, presents significant challenges due to the variability of digital learning environments in education. While generic process indicators can lead to ambiguous interpretations, process indicators derived from domain-specific theories can reveal meaningful insights beyond mere outcome measures.

To summarize, log data offer a powerful way to capture relevant indicators in teachers' assessment processes, provided that they are recorded in well-designed simulations that allow for theoretically grounded interpretations.

5. The present research

The starting point for this dissertation project was the need for evidence-based and practice-oriented training of mathematics teachers' core practices at the university level. It is argued that engaging in practical challenges enables pre-service teachers to apply their theoretical knowledge in their later classroom practice, which is likely to lead to higher learning gains in students.

The present research focuses on teachers' assessments of students' competencies and misconceptions as one emblematic core practice in mathematics teaching. To provide practice-oriented training, we used a digital simulation of an interactive assessment situation. In the simulated situation, pre-service teachers had to assess individual students' mathematical competencies and misconceptions. We investigated teachers' processes of selecting tasks and interpreting student solutions while pre-service teachers were engaged in the assessment situations. Going beyond previous research that often solely focused on the accuracy of assessments, a main goal was to analyze how pre-service teachers' dispositions and assessment process contributed to accurate assessments. This involved implementing different types of scaffolding for supporting pre-service teachers' assessment process in the simulation. In addition, we used log data from the digital simulation to analyze individual differences in pre-service teachers' assessment process. These analyses should contribute to the implementation of adaptive scaffolding that is tailored to individual differences further enhance teachers' assessment performance. In summary, the overarching question that guided the present research was:

RQ0: How can we effectively enhance mathematics teachers' assessment accuracy?

Study A aimed to evaluate both the design of the simulation of an interactive assessment situation and the effects of different types of scaffolding on assessment accuracy. To this end, we first compared learners' experience while engaging in different versions of the simulation. The study examined how different types of scaffolding during the assessment process influenced assessment accuracy.

The specific research questions were:

RQ1a: Does the implementation of different types of scaffolding in the simulation affect individuals' perceived presence, authenticity, and extraneous cognitive load, relative to the simulation without scaffolding?

RQ1b: Does using the simulation increase prospective primary school teachers' accuracy regarding students' competence levels and their specific misconceptions?

RQ1c: Does scaffolding (either content-related or strategic) have a positive effect on assessment accuracy regarding students' mathematical competencies and misconceptions? Is content-related scaffolding more effective than strategic scaffolding?

Building on Study A, we aimed to investigate in more detail how learners differ in the quality of their assessment processes, and how these differences are related to dispositions and assessment accuracy. Accordingly, Study B is guided by the question of how scaffolding, as used in Study A, can be adapted to individual learners' needs. To this end, the assessment processes were captured using log data from the digital simulation. Additionally, individual differences were identified and analyzed in terms of how they might inform the implementation of adaptive support. The specific research questions for Study B were:

RQ2a: Can we reliably measure pre-service teachers' situation-specific indicators in a person-related assessment situation across multiple assessment items using log data from a digital simulation?

RQ2b: Are there individual differences in pre-service teachers' assessment processes based on their patterns of diagnostic activities in a digital simulation?

RQ2c: How are pre-service teachers' cognitive and affective-motivational dispositions and their assessment process related to each other and to their accuracy?

In addition to these two studies, more specific analyses and preliminary work were published in conference proceedings. This work also contributed to the overarching research goal of this project (Section 6.3).

6. Summary of publications

6.1 Evaluation of design and effectiveness of a digital simulation with scaffolding (Paper A)

Schons, C., Obersteiner, A., Reinhold, F., Fischer, F., & Reiss, K. (2023). Developing a simulation to foster prospective mathematics teachers' diagnostic competencies: the effects of scaffolding. *Journal für Mathematik-Didaktik*, *44*(1), 59–82. https://doi.org/10.1007/s13138-022-00210-0

6.1.1 Research goals

The overarching goal of Paper A was to evaluate the design and effectiveness of simulations with different types of scaffolding. Despite their potential for acquiring professional competencies, simulations have rarely been used for mathematics teacher training. Building on mathematics teachers' core practices, we used a digital simulation of an assessment situation in which pre-service teachers had to assess students' mathematical competencies and misconceptions. We implemented two different types of scaffolding (strategic and content-related scaffolding) to support the assessment process in the simulation.

The first goal of Study A was to evaluate whether the different types of scaffolding affected learners' experience in the simulation. Thus, we compared learners' experience while using different versions of the simulation in terms of their perceived presence, authenticity, and extraneous cognitive load (RQ1a). The second goal was to evaluate the extent to which strategic and content-related scaffolding during the assessment process can enhance pre-service teachers' assessment accuracy. To this end, we first compared the effectiveness of the simulation with and without scaffolding (RQ1b). In addition, we investigated whether content-related support was more effective than strategic support in fostering pre-service teachers' assessment competencies (RQ1c).

6.1.2 Methodology

The digital simulation. The simulation was developed using the CASUS e-learning software of the non-profit company INSTRUCT gGmbH as a platform. This software provides a framework for creating authentic assessment tasks. An initial version of the simulation for analyzing and fostering mathematics teachers' assessment competencies is described in Wildgans-Lang et al. (2020). This simulation served as the basis for Study A and was thoroughly revised. In addition, strategic and content-related scaffolding were implemented into the simulation.

In the simulation, pre-service teachers assess simulated students' mathematical competencies by selecting tasks and interpreting the simulated students' written solutions. Pre-service teachers could select blank tasks from a provided portfolio to gather information about a simulated student's competencies and misconceptions. Upon selecting a task, the simulated student's solution was immediately displayed. The order in which the tasks were selected was determined by the participants' individual choices, and they could stop selecting tasks at any time to complete the assessment. A key feature of the simulation was that the tasks and solutions were taken from the national large-scale assessment among third-graders in Germany (VERA-3; "Vergleichsarbeiten"). Consequently, the tasks were aligned to one of five competence levels according to Reiss and Winkelmann's (2009) validated competence model. The task solutions came from real students who participated in the VERA-3 pilot study.

To construct the simulated students' task portfolios, we selected written task solutions from the VERA-3 item pool that showed specific competencies and misconceptions. Each simulated student was assigned a common misconception (e.g., regarding the place-value system) that is well documented in the mathematics education literature (Padberg & Benz, 2021). Similarly, students' competence levels could be assigned to one of five levels according to the competence model based on the simulated students' solution rates. The assessment goal for the participants in Study A was to match each simulated student to a competence level and a misconception.

Scaffolding. The content-related scaffolding provided detailed information about the competence model that served as the reference for the tasks in the simulation. When participants in this condition selected a task from the portfolio, descriptions of the competence levels were displayed during task selection. These descriptions remained permanently visible until the interpretation of the task solution was completed. This type of scaffolding was meant to serve pre-service teachers as a reference norm, supporting the application of knowledge during the interpretation of student solutions.

The strategic scaffolding included information about essential activities during the assessment process. It was based on Heitzmann et al.'s (2019) and Fischer et al.'s (2014) models of activities in scientific reasoning (Section 2.1.2). The strategic scaffolding was displayed in the same location on the screen as the content-related scaffolding and was continuously visible during the assessment process. Strategic scaffolding was meant to support pre-service teachers' quality of reasoning during the interpretation of student solutions.

Study design. Participants in Study A were pre-service teachers from two German universities and were assigned to four conditions. The first condition was the control condition (N = 196), in which participants assessed two simulated students in the simulation without any intervention. These two simulated students served as the pretest and posttest for all conditions. Participants in the second condition (N = 20) received an intervention with the simulation, but without scaffolding. Participants in the third (N = 24) and fourth (N = 18) condition received an intervention with content-related or strategic scaffolding, respectively. After assessing simulated students in the simulation, participants in the intervention conditions were asked about their perceived presence, authenticity, and cognitive load using established scales.

6.1.3 Main findings

Participants in all conditions rated their perceived presence and authenticity as rather high and their perceived cognitive load as moderately high, indicating that they were engaged in the assessment situation without experiencing cognitive overload. Differences between the conditions were small and not statistically relevant, indicating that the implementation of scaffolding had only negligible effects on pre-service teachers' experience in the simulation.

Participants in the intervention conditions were 62 % more likely to make an accurate assessment of the competence level on the posttest compared to the control condition. However, there was no statistically significant difference in assessing students' misconceptions between the intervention and control conditions. Furthermore, descriptive data indicated that the participants who received content-related scaffolding were 13 % more likely to select the correct competence level and 83 % more likely to identify the misconception accurately in the posttest compared to participants who received strategic scaffolding. Although these differences were not statistically significant, the results suggest that content-related scaffolding had a positive effect on identifying students' misconceptions accurately compared to strategic scaffolding.

6.2 Identification of individual differences in the assessment process and associations with dispositions and assessment accuracy (Paper B)

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2024). Toward adaptive support of pre-service teachers' assessment competencies: Log data in a digital simulation reveal engagement modes. *Learning and Instruction*, *94*, 101979. https://doi.org/10.1016/j.learninstruc.2024.101979

6.2.1 Research goals

Study B built on the findings of Study A, and the overarching goal was to further improve the effectiveness of scaffolding in the simulation for acquiring assessment competencies. To achieve this, it was necessary to improve our understanding of the relationships between dispositions and the assessment process during the genesis of accurate assessments. Accordingly, instead of indirectly manipulating the assessment process through different types of scaffolding, we aimed to measure the assessment process directly through process indicators that could be captured through the log data of the digital simulation. In this way, individual differences in pre-service teachers' assessment process could be analyzed for developing adaptive scaffolding.

The first goal of Study B was to ensure that the process indicators could be reliably measured through the log data of the digital simulation (RQ2a). Furthermore, we were interested in individual differences in the quality of the assessment process among pre-service teachers (RQ2b) and how these were related to dispositions and assessment accuracy (RQ2c). Thus, the interplay between dispositions and the assessment process may provide insights into whether they serve as two distinct leverage points for adaptation of scaffolding or whether both are strongly related and influence assessment performance.

6.2.2 Methodology

The digital simulation. The simulation (without scaffolding) from Study A served as the basis for this study and we retained the core functionality of the simulation. However, we revised the simulated students in order to capture the assessment process through log data. This allowed us to use process indicators that covered both the process of selecting tasks and the process of interpreting task solutions. The simulated students' task portfolios in the simulation were carefully revised so that each contained a fixed number of tasks with diagnostic potential to reveal a specific misconception (Figure 5). The reduction to a single assessment goal (focus on misconceptions, not competency levels) ensured a reliable coding of the notes pre-service teachers took during the interpretation of student solutions.

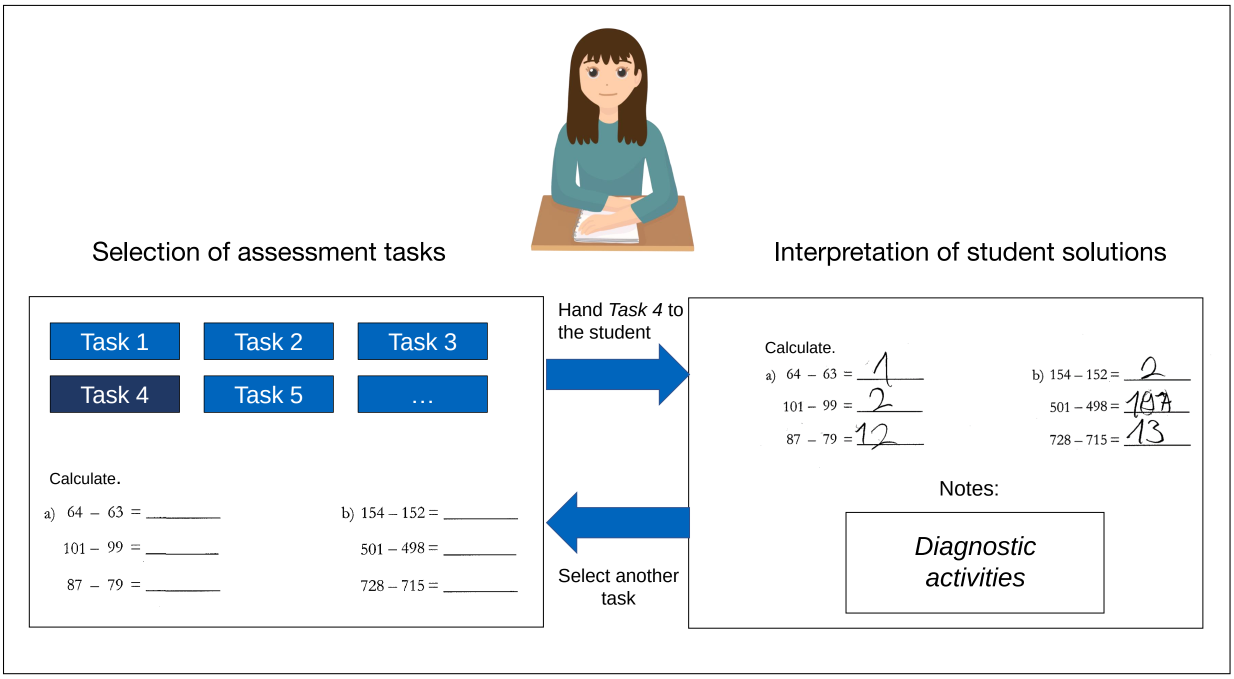

Figure 5 Illustration of the assessment of an individual student in the simulation. The process of selecting tasks was captured through the click data on tasks, the interpretation of student solutions was captured through the written notes.

Study design. Participants in Study B were $N = 65$ pre-service mathematics teachers studying at a German university. Before engaging in the simulation, participants' cognitive (CK and PCK) and affective-motivational (individual interest and self-efficacy) dispositions were measured. Afterwards, participants assessed up to eight simulated students in the digital simulation (Figure 6). In Study B, all participants worked in the same condition in the digital simulation.

6.2.3 Main findings

First, pre-service teachers' process indicators were measured with high reliability using log data from the digital simulation. Additionally, we identified three distinct patterns of pro-

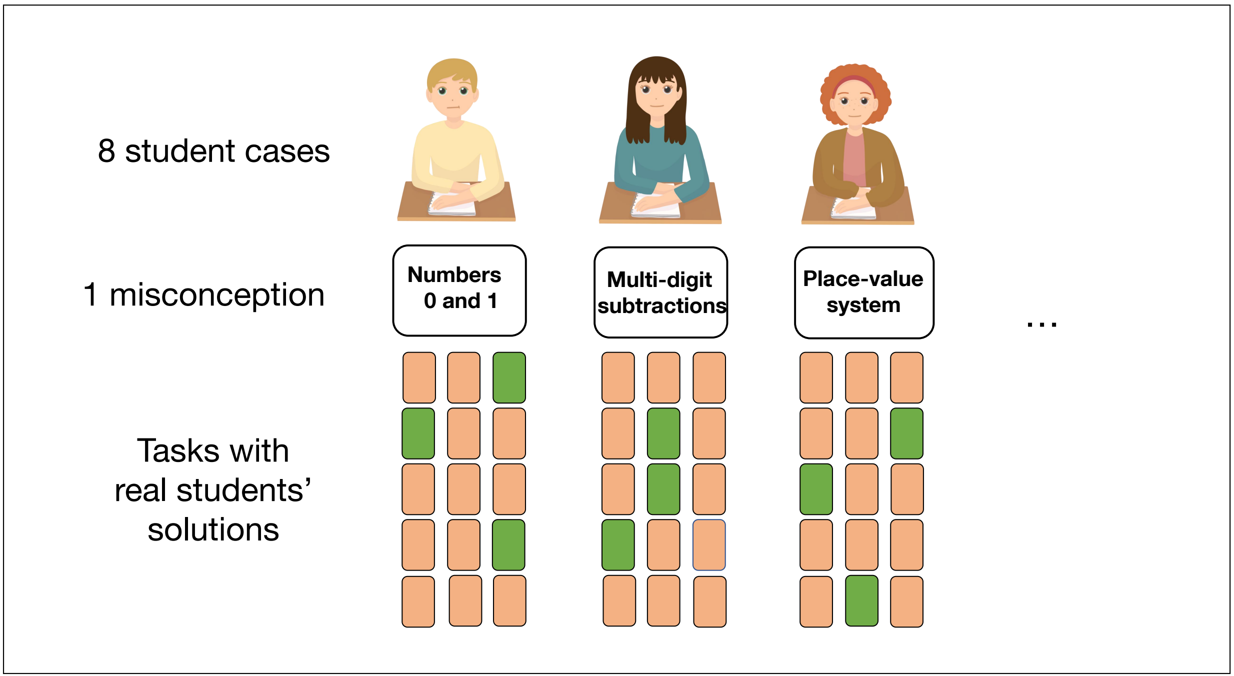

Figure 6 Overview of the simulated students in the simulation used in Study B. Participants had to select and interpret tasks to assess a mathematical misconceptions. The tasks either had (green) or did not have (orange) diagnostic potential to reveal an error related to the misconception.

cess indicators in our sample of pre-service teachers, revealing qualitative differences in pre-service teachers' assessment processes. The first pattern reflected a passive mode of cognitive engagement, characterized by selecting a below-average number of tasks and writing a below-average number of notes while interpreting student solutions. The second pattern indicated an active mode by selecting an above-average number of tasks and writing an above-average number of notes, but without considering the diagnostic potential of the tasks and primarily writing superficial descriptions of task solutions. The third pattern suggested a constructive engagement mode, in which participants carefully selected tasks and generated an above-average number of hypotheses when interpreting the task solutions.

Additional analyses showed that pre-service teachers with constructive engagement had significantly more interest in student assessment than pre-service teachers with passive engagement, while other affective-motivational dispositions and professional knowledge did not differ significantly between engagement modes.

Pre-service teachers' mathematical content knowledge was related to assessment accuracy. Above and beyond professional knowledge, the different patterns of diagnostic activities during the assessment process predicted assessment accuracy: Pre-service teachers with constructive engagement assessed the virtual students' misconceptions significantly more accurately than pre-service teachers with passive engagement.

6.3 Associated publications

In addition to the two main publications, there are conference papers related to this project which are briefly summarized in this section.

Two papers were published in the proceedings of the annual meeting of the Society for Didactics of Mathematics (GDM). In Schons et al. (2020), the simulation was validated after the implementation of scaffolding as part of the preparation for Study A. Schons, Obersteiner, Fischer, and Reiss (2023) analyzed the distribution of diagnostic activities during pre-service teachers' interpretation of student solutions. A key finding was that pre-service teachers focused predominantly on evaluating evidence by describing superficial features of task solutions, and hardly generated hypotheses or drew conclusions.

Two additional contributions were published in the proceedings of the annual meeting of the International Group for the Psychology of Mathematics Education (PME). Schons, Obersteiner, Fischer, and Reiss (2022) analyzed the task selection process. Specifically, we analyzed whether participants considered the diagnostic potential of the tasks. Using sequence clustering (Gabadinho et al., 2011), which considers patterns of activities including the temporal order of task selection, we identified distinct profiles: In one profile, pre-service teachers selected almost all available tasks. In another, they initially selected many tasks without diagnostic potential before eventually selecting those with diagnostic potential. In a third profil, pre-service teachers had a more deliberate procedure by selecting fewer tasks overall but a significant number of tasks with diagnostic potential. These findings underscored substantial individual differences in the quality of task selection during the assessment process.

Schons, Obersteiner, Fischer, and Reiss (2023) evaluated the coding scheme for categorizing written notes into the activities during the interpretation of task solutions (evaluating evidence, generating hypotheses, and drawing conclusions) in terms of inter-coder reliability. Additionally, we employed a person-centered approach to examine individual differences in interpreting student solutions. The results revealed that one group of pre-service teachers focused primarily on evaluating evidence, with minimal engagement in other diagnostic activities. In contrast, a smaller group also generated hypotheses and drew conclusions. These findings highlight significant individual differences in the quality of interpreting student solutions. The results suggest that a person-centered approach is beneficial for developing adaptive support.

7. Discussion

This dissertation project investigated how to foster the assessment competencies of pre-service mathematics with a digital simulation. The project aimed to provide evidence-based practice opportunities for pre-service teachers during their university education. Such practice opportunities aim to facilitate the transfer of theoretical knowledge into practical application, ensuring that pre-service teachers are well prepared for real classroom situations. The studies presented in this dissertation highlight the potential of simulations for teacher training and offer novel methodological possibilities that can guide future research on mathematics teachers' professional competencies.

The two studies focused on mathematics teachers' assessment competencies as an emblematic core practice of mathematics teachers. In particular, we focused on the processes of selecting tasks and interpreting student solutions in assessment situations with individual students. Pre-service teachers had to assess simulated students' mathematical competencies and misconceptions.

First, a digital simulation with different types of scaffolding was developed and evaluated in terms of design and effectiveness for fostering pre-service teachers' assessment accuracy. With the aim of implementing adaptive scaffolding in the simulation and thus improving the effectiveness of the simulation, we measured the assessment process holistically using log data. Using log data from a digital simulation allowed for a comprehensive measurement of pre-service teachers' assessment competencies, including dispositions, the assessment process, and assessment performance.

7.1 Design and effectiveness of simulations for training mathematics teachers' assessment competencies

Simulations have proven effective across various domains (Chernikova, Heitzmann, Stadler, et al., 2020), but have rarely been used in teacher education. In this project, a simulation was developed to provide a practice opportunity for pre-service mathematics teachers. The simulation was developed based on general design principles to create an assessment situation that is relevant to mathematics teachers (Wildgans-Lang et al., 2020). To ensure its authenticity, the situation was designed based on mathematics educational considerations.

7.1.1 Design of the simulation

Pre-service teachers rated the simulation as authentic and immersive. Together with the finding that the cognitive load was perceived as moderately high, it can be concluded that an appropriate balance between authenticity and cognitive demand was achieved in the simulation used in this project (Codreanu et al., 2020).

Pre-service teachers' learning experience did not substantially differ between simulations with different types of scaffolding or without scaffolding, and the situation was perceived as authentic in all conditions. This result suggests that integrating scaffolding into simulations does not necessarily influence learners' perception of the simulation and offers comparable conditions regarding the overall learning experience.

The simulation represented a person-related assessment situation in which pre-service teachers had to assess students' mathematical competencies and misconceptions based on their task solutions. Although a one-to-one interview situation is not a typical example of current classroom practice, the simulation was rated as authentic by pre-service teachers. This suggests that prospective teachers do not perceive the situation as unnatural and are willing to immerse themselves in such situations. Given the call for more student-centered teaching (Weimer, 2013), this is encouraging since pre-service teachers are willing to master the assessment of individual students.

The simulation used a case-based scenario, which is commonly used in medical simulations (McLean, 2016). In medicine, it has proven to be an effective method of fostering assessment competencies through cases that include a wide range of assessment challenges and are designed in accordance with medical knowledge of common diseases. We transferred the case-based learning approach to mathematics teacher training. Compared to the field of medicine, in teacher education it is often less clear which cues indicate which competencies when assessing mathematical competencies. Therefore, a theoretically grounded and empirically validated competence model was used as the basis for the simulation and served as an objective reference for the assessments. The need for an objective reference standard has often been neglected in previous research, yet it is a crucial foundation for a systematic improvement of teachers' assessment competencies (Südkamp et al., 2012).

7.1.2 Effectiveness of the simulation with scaffolding

Building on research on fostering complex skills, it was expected that working with the simulation would have a large effect on pre-service teachers' assessment performance. However, the results of Study A suggest that working with the simulation alone does not affect assessment performance when operationalized as accuracy. The simulation only had an effect on assessment accuracy when additional instructional support was provided in the form of content-related scaffolding. The content-specific scaffolding included descriptions of levels of mathematical competence levels and was designed to activate their pedagogical content knowledge. At the same time, the competence model also served as a reference for making assessments. Thus, the positive effect aligns with the finding that teachers make more accurate assessments when they are informed about the specific criteria used to assess the students (Südkamp et al., 2012).

Strategic scaffolding, however, had no effect on pre-service teachers' accuracy. This could be due to the limited experience of pre-service teachers in this kind of assessments. The strategic scaffolding did not include relevant content for the simulation but rather focused on the assessment process, specifically guiding pre-service teachers to use a hypothesis-driven strategy. Scaffolding that provides support on a meta-level rather than referring to specific content of the learning task tends to be more effective for experienced learners. The sample in Study A might not have had the necessary knowledge and experience to benefit from this type of scaffolding (Chernikova, Heitzmann, Stadler, et al., 2020). However, we can only speculate about this because the study did not measure pre-service teachers' professional knowledge, which is a limitation of Study A. Another limitation is that activities such as stating hypotheses or drawing conclusions in the assessment process were not measured. Therefore, we cannot determine whether strategic scaffolding at least had an effect on pre-service teachers' diagnostic activities during the assessment process, even if it did not have an effect on accuracy.

Similar studies have also found that short interventions – even adaptive interventions – tend to have only minor effects on assessment accuracy (Nickl, Sommerhoff, Böheim, et al., 2024; Nickl, Sommerhoff, Radkowitsch, et al., 2024; Sommerhoff et al., 2023). The results of Study A align with these findings. Improving the effectiveness of scaffolding during the assessment process requires a more detailed analysis of the specific activities that contribute to accurate assessments. By identifying which specific activities in the assessment process are relevant for accurate assessments, more tailored scaffolding can be developed, which in turn should be more effective in enhancing accuracy than static scaffolding such as those used in Study A.

7.2 The relationship between dispositions and the assessment process in the genesis of assessments

Previous research focusing on mathematics teachers' assessment processes has often examined these processes in highly controlled settings through experimental variation (Brunner et al., 2024; Ostermann et al., 2018; Rieu et al., 2022). In this project, we investigated the assessment process in a situation including teacher-student interaction by using multiple process indicators derived from log data. The indicators refer to both the processes of selecting tasks and interpreting student solutions. This approach allowed us to capture the assessment process holistically and to identify individual differences that can be used to develop adaptive scaffolding.

7.2.1 Use of log data to analyze the assessment process

The overarching question in this research project was how to effectively enhance the mathematics teachers' assessment accuracy. Study B focused on how pre-service teachers' dispositions and the assessment process contribute to the genesis of accurate assessments. The operationalization of the assessment process was challenging, as it involved an interactive assessment situation in which the process was navigated by the pre-service teachers themselves. Under these conditions, the log data from the digital simulation provided a unique and unobtrusive way of capturing individual processes.

To account for the interaction between pre-service teachers and simulated students in the digital simulation, multiple process indicators were used to indicate both task selection and the interpretation of task solutions. This holistic approach for capturing the assessment process acknowledges that the quality of the process is not determined by analyzing individual process indicators in isolation, but rather by the interaction between them. This was achieved by identifying patterns in the configurations of process indicators. This approach went beyond previous studies that operationalized the assessment process through single indicators or examined them in isolation.

Furthermore, the identified patterns were interpreted within a theoretical framework, the ICAP framework. This framework describes qualitative differences in learning behavior in interactive situations (Chi & Wylie, 2014). While the ICAP framework has been criticized for only superficially describing learning behavior (Thurn et al., 2023), the process indicators in Study B were explicitly formulated for the context of an interactive assessment situation, thereby explicitly and deeply describing the engagement modes. This approach can serve as an example of how the ICAP framework can be used effectively to identify qualitative differences in learning processes in complex learning situations.

7.2.2 Development of adaptive scaffolding

The findings of Study B can be used to implement adaptive scaffolding with the aim of improving pre-service teachers' assessment accuracy. The study examined the extent to which cognitive and affective-motivational dispositions, as well as their assessment process, contribute to the genesis of accurate assessments. Among the dispositions, professional knowledge was directly related to assessment accuracy — consistent with studies on task-related assessments (Brunner et al., 2024; Ostermann et al., 2018; Rieu et al., 2022) and person-related assessments (Kron et al., 2022).

As expected, affective-motivational dispositions were not directly related to accuracy. However, individual interest in student assessment was linked to the quality of the assessment process, which, in turn, was related with assessment accuracy. A key finding was that these differences in the assessment process were not explained by professional knowledge, making professional knowledge and the assessment process two distinct leverage points for fostering assessment competencies. While direct instruction of professional knowledge has proven effective in enhancing pre-service teachers' accuracy (Rieu et al., 2022), effective scaffolding during the assessment process holds potential to further enhance assessment accuracy. Individual differences in the diagnostic processes were identified, suggesting that pre-service teachers may require different types of scaffolding.

Study B established the basic steps for implementing adaptive scaffolding in line with Plass and Pawar's (2020) framework. Combined with the findings from Study A, particularly the finding that content-related scaffolding had a positive effect on accuracy, the following adaptive scaffolding mechanism can be proposed (Figure 7): After the assessment of simulated student cases, the process indicators are analyzed and form the adjustment basis for allocating pre-service teachers to the constructive, active or passive engagement mode by applying a clustering algorithm. Pre-service teachers in the constructive mode could receive content-specific scaffolding to support their task selection, while those in the active mode could additionally benefit from prompts about the misconceptions to be assessed. Pre-service teachers in the passive mode, characterized by low interest in student interest, may benefit from motivational scaffolding.

Future studies could implement adaptive scaffolding of the assessment process to investigate whether it can indeed enhance assessment accuracy beyond direct instruction of professional knowledge prior to the assessment situation.

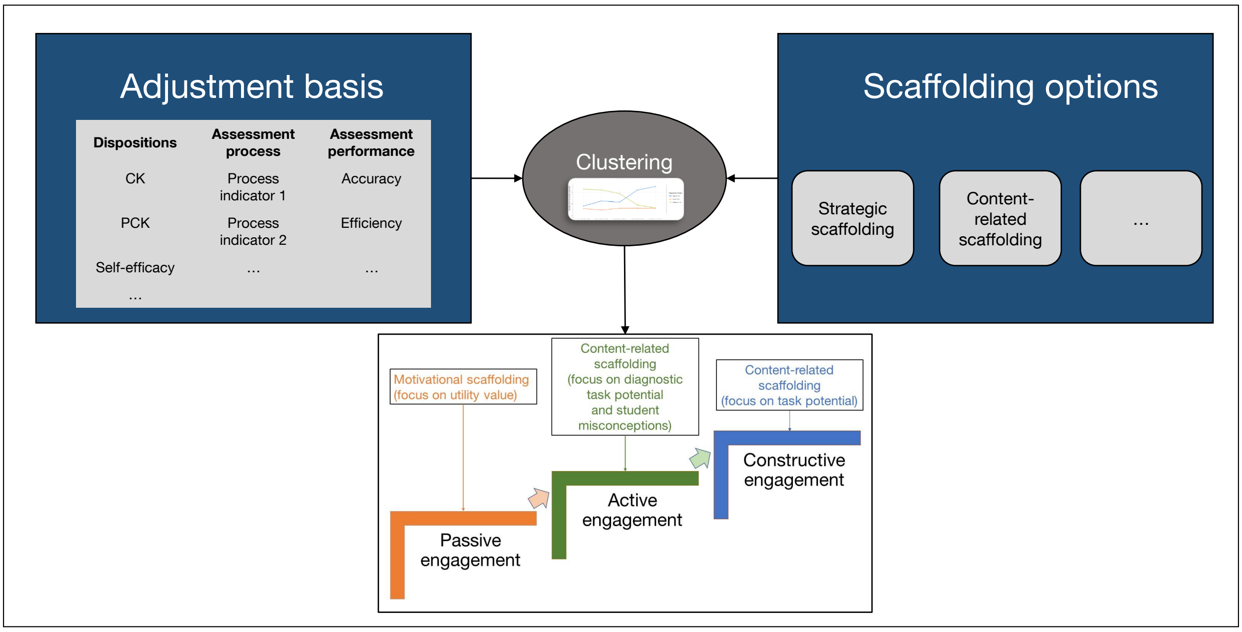

Figure 7 Overview of an adaptivity mechanism: Pre-service teachers are assigned specific engagement modes based on multiple process indicators. Depending on the engagement mode, targeted scaffolding is provided.

7.3 Methodological considerations

Log data offer potential for research, but require a theoretical foundation for the interpretation. Due to the complexity of log data, a meaningful analysis also requires advanced statistical methods. In this project, theoretically grounded process indicators for an interactive assessment situation were defined and captured using the log data from a digital simulation. For the statistical analysis, cluster analysis was initially performed on these process indicators, followed by an analysis of how the resulting cluster and professional knowledge predicted assessment accuracy. The strength of this approach has already been demonstrated in studies that used process data in educational research (Radkowitsch et al., 2023; Reinhold et al., 2020). This approach is particularly effective for more complex learning or assessment processes, where the interaction between process indicators is more important than individual indicators, as was the case in this project.

During the course of this dissertation project, several methodological challenges emerged that may also be relevant for future research on digital learning environments and simulations. Study B demonstrated how log data can be used to analyze activities during the assessment process of pre-service mathematics teachers. The activities were aggregated over time. However, a further step could be to consider the temporal order in which these indicators occur, such as in sequence analysis (e.g., Schons et al., 2023). A key challenge is to provide a theoretically grounded interpretation of such sequences, as existing models often do not offer hypotheses about the temporal patterns of activities. Nevertheless, the ability of log data to capture and analyze the temporal dimension represents a research potential for analyzing assessment processes, although it requires further development of current theoretical models to fully realize this potential.

In Study B, the analysis of consistency was a prerequisite for ensuring that the aggregation of process indicators across multiple student assessments could reveal individual differences. We realized that studies using process indicators often did not report the consistency across items, even though the process indicators were aggregated for further analyses. However, this makes it challenging to derive adaptive support from these analyses. It remains unclear whether the process indicators are stable within individual learners or within items. As a result, it remains unclear whether the adaptivity mechanism should be based on differences between individual learners (if the process indicators are stable within individual learners) or on differences in the characteristics of the test items (if the process indicators are stable within test items). For the further development of adaptivity mechanisms, it would therefore be desirable for future studies that analyze process indicators to also report the consistency of the indicators across items.

7.4 Limitations and future directions

Beyond the limitations reported in the published Papers A and B, this project also had at least two limitations that could be addressed in future research on simulations for fostering teachers' professional competencies. First, some of the errors that served as indicators of student misconceptions in the simulation were procedural rather than conceptual. This limitation was due to the decision to use the competence model as the reference norm for the assessments, which also determined the tasks. Future research on assessing misconceptions could address this limitation by focusing more on conceptual errors related to student misconceptions and then developing tasks specifically for the purpose of assessing individual students.

Second, the analysis and promotion of assessment competencies in this project were limited to very short time periods, during which only a small number of simulated students were assessed, and only one specific mathematical topic was covered. However, it is likely that the professional competencies of pre-service teachers develop over a longer period, supported by mathematics education courses. However, to integrate case-based simulations into university teaching to develop assessment competencies over longer periods and across various topics, more knowledge about individual students' misconceptions is needed. Currently, it is well known how prevalent certain mathematical misconceptions are across populations, but there is little knowledge about how consistently individual students make specific errors or which combinations of errors occur in individual students. This knowledge, however, is essential for simulating authentic students. In medical education, there is a large body of such knowledge and a long tradition of case-based learning, and medical education can serve as a model for further implementing the case-based learning approach with simulations in mathematics teacher education.

With regard to the further development of adaptive scaffolding, a future direction is the analysis of process indicators in real time. The results from this project can be used to inform decisions about what scaffolding might be beneficial based on specific process patterns.

However, all analyses were conducted post-hoc rather than in real time. A key advancement of evaluating these process indicators in real time would be the ability to implement micro-adaptive scaffolding. Beyond cognitive process indicators as used in this project, affective-motivational or emotional indicators could also be included in the real time adjustment basis (Plass and Pawar, 2020). A significant methodological challenge here is how to validly and reliably assess non-cognitive indicators in real time (Foulkes et al., 2023).

8. Conclusion

Assessing students' mathematical competencies and misconceptions is one of the core practices of mathematics teachers. The research project presented here aimed to provide an evidence-based practice opportunity for pre-service teachers with the goal to effectively foster their assessment competencies during their university education.

A simulation providing authentic simulated students was used to holistically analyze and foster pre-service mathematics teachers' assessment competencies. The design of the simulation allowed for reliable measurement of multiple process indicators to capture the assessment process in an interactive assessment situation, including both task selection and the interpretation of student solutions. This approach allowed us to go beyond simply describing assessment accuracy by investigating how dispositions and the assessment process contribute to the genesis of accurate assessments. Based on these findings, we derived recommendations for adaptive scaffolding that can be applied in future research.

Overall, this research project can guide the development and evaluation of simulations aimed at fostering mathematics teachers' core practices, and thus contributing to evidence-based and more practice-oriented teacher training in mathematics education.

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

Anders, Y., Kunter, M., Brunner, M., Krauss, S., & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, *57*(3), 175–193. https://doi.org/10.2378/peu2010.art13d

Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, *23*(34), 157–160. https://doi.org/10.1024/1010-0652.23.34.157

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 427–449, Vol. 28). Springer New York. https://doi.org/10.1007/978-1-4419-5546-3_28

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bardy, T., Holzäpfel, L., Reinhold, F., & Leuders, T. (2024). Mathematics teachers' multiple perspectives on adaptive tasks: task evaluation and selection as core practices for teaching quality. *ZDM – Mathematics Education*. https://doi.org/10.1007/s11858-024-01626-3

Bauer, E., Fischer, F., Kiesewetter, J., Shaffer, D. W., Fischer, M. R., Zottmann, J. M., & Sailer, M. (2020). Diagnostic activities and diagnostic practices in medical education and teacher education: an interdisciplinary comparison. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.562665

Behrmann, L., & Souvignier, E. (2013). The relation between teachers' diagnostic sensitivity, their instructional activities, and their students' achievement gains in reading. *Zeitschrift für Pädagogische Psychologie*, *27*(4), 283–293. https://doi.org/10.1024/1010-0652/a000112

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

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. J. (2015). Beyond dichotomies. *Zeitschrift für Psychologie*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Brom, C., Děchtěrenko, F., Frollová, N., Stárková, T., Bromová, E., & D'Mello, S. K. (2017). Enjoyment or involvement? Affective-motivational mediation during learning from a complex computerized simulation. *Computers & Education*, *114*, 236–254. https://doi.org/10.1016/j.compedu.2017.07.001

Brunner, K., Obersteiner, A., & Leuders, T. (2024). How pedagogical content knowledge sharpens prospective teachers' focus when judging mathematical tasks: an eye-tracking study. *Educational Studies in Mathematics*, *115*(2), 177–196. https://doi.org/10.1007/s10649-023-10281-6

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

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

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

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

Codreanu, E., Sommerhoff, D., Huber, S., Ufer, S., & Seidel, T. (2021). Exploring the Process of Preservice Teachers' Diagnostic Activities in a Video-Based Simulation. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.626666

Cook, D. A., Brydges, R., Hamstra, S. J., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hatala, R. (2012). Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis. *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, *7*(5), 308–320. https://doi.org/10.1097/SIH.0b013e3182614f95

Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical Teacher*, *35*(1). https://doi.org/10.3109/0142159X.2012.714886

Corno, L. (2008). On teaching adaptively. *Educational Psychologist*, *43*(3), 161–173. https://doi.org/10.1080/00461520802178466

Cotton, D., & Gresty, K. (2006). Reflecting on the think–aloud method for evaluating e-learning. *British Journal of Educational Technology*, *37*(1), 45–54. https://doi.org/10.1111/j.1467-8535.2005.00521.x

Dawley, L., & Dede, C. (2014). Situated learning in virtual worlds and immersive simulations. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 723–734). Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_58

de Corte, E., Verschaffel, L., & Schrooten, H. (1991). Computer simulation as a tool in studying teachers' cognitive activities during error disgnosis in arithmetic. In P. Goodyear (Ed.), *Teaching knowledge and intelligent tutoring* (pp. 367–378). Intellect Books.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, *68*(2), 179–201. https://doi.org/10.3102/00346543068002179

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

Eichelmann, A., Narciss, S., Schnaubert, L., & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen – Ein Review zu empirischen Fehleranalysen. *Journal für Mathematik-Didaktik*, *33*(1), 29–57. https://doi.org/10.1007/s13138-011-0031-5

Fanning, R. M., & Gaba, D. M. (2007). The role of debriefing in simulation-based learning. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, *2*(2), 115–125. https://doi.org/10.1097/SIH.0b013e3180315539

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304. https://doi.org/10.1504/IJTEL.2012.051816

Fischer, F., Bauer, E., Seidel, T., et al. (2022). Representational scaffolding in digital simulations – learning professional practices in higher education. *Information and Learning Sciences*, *123*(11/12), 645–665. https://doi.org/10.1108/ILS-06-2022-0076

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *2*(3), 28–45. https://eric.ed.gov/?id=EJ1090940

Fouh, E., Breakiron, D. A., Hamouda, S., Farghally, M. F., & Shaffer, C. A. (2014). Exploring students learning behavior with an interactive e-textbook in computer science courses. *Computers in Human Behavior*, *41*, 478–485. https://doi.org/10.1016/j.chb.2014.09.061

Foulkes, M., Weiers, H., & Schons, C. (2023). Investigating the relationship between state mathematics anxiety and performance using physiological real-time measures. https://doi.org/10.17605/OSF.IO/BT9J4

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4). https://doi.org/10.18637/jss.v040.i04

Goldhammer, F., Hahnel, C., Kroehne, U., & Zehner, F. (2021). From byproduct to design factor: on validating the interpretation of process indicators based on log data. *Large-scale Assessments in Education*, *9*(1), 1–25. https://doi.org/10.1186/s40536-021-00113-5

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608–626. https://doi.org/10.1037/a0034716

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, *45*(4), 203–209. https://doi.org/10.1080/00461520.2010.515935

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior*, *61*, 36–46. https://doi.org/10.1016/j.chb.2016.02.095

Grossman, P. L., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record: The Voice of Scholarship in Education*, *111*(9), 2055–2100. https://doi.org/10.1177/016146810911100905

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

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers*, *3*, 275–285. https://doi.org/10.1016/j.susoc.2022.05.004

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

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*(4), 1–24. https://eric.ed.gov/?id=EJ1232563

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

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, *42*(2), 371–406. https://doi.org/10.3102/00028312042002371

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, *59*(3), 297–313. https://doi.org/10.3102/00346543059003297

Kolovou, D., Hochweber, J., & Praetorius, A.-K. (2024). Does teacher judgment accuracy matter? How judgment accuracy, teaching quality, and student achievement development are related. *Teaching and Teacher Education*, *144*, 104555. https://doi.org/10.1016/j.tate.2024.104555

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2016). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110. https://doi.org/10.18608/jla.2015.23.6

Kron, S., Sommerhoff, D., Achtner, M., Stürmer, K., Wecker, C., Siebeck, M., & Ufer, S. (2022). Cognitive and motivational person characteristics as predictors of diagnostic performance:

Combined effects on pre-Service teachers' diagnostic task selection and accuracy. *Journal für Mathematik-Didaktik*, *43*(1), 135–172. https://doi.org/10.1007/s13138-022-00200-2

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

Kultusministerkonferenz. (2004). Standards der Lehrerbildung: Bildungswissenschaften [Standards of teacher education: Educational sciences].*Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany, Berlin*.

Larrain, M., & Kaiser, G. (2022). Interpretation of Students' Errors as Part of the Diagnostic Competence of Pre-Service Primary School Teachers. *Journal für Mathematik-Didaktik*, *43*(1), 39–66. https://doi.org/10.1007/s13138-022-00198-7

Leuders, T., Loibl, K., Sommerhoff, D., Herppich, S., & Praetorius, A.-K. (2022). Toward an overarching framework for systematizing research perspectives on diagnostic thinking and practice. *Journal für Mathematik-Didaktik*, *43*(1), 13–38. https://doi.org/10.1007/s13138-022-00199-6

Levin, O., Frei-Landau, R., Flavian, H., & Miller, E. C. (2023). Creating authenticity in simulation-based learning scenarios in teacher education. *European Journal of Teacher Education*, 1–22. https://doi.org/10.1080/02619768.2023.2175664

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

McLean, S. F. (2016). Case-based learning and its application in medical and health-care fields: a review of worldwide literature. *Journal of Medical Education and Curricular Development*, *3*. https://doi.org/10.4137/JMECD.S20377

Nickl, M., Sommerhoff, D., Böheim, R., Ufer, S., & Seidel, T. (2024). Fostering pre-service teachers' assessment skills in a video simulation. *Zeitschrift für Pädagogische Psychologie*, *38*(1-2), 27–34. https://doi.org/10.1024/1010-0652/a000362

Nickl, M., Sommerhoff, D., Radkowitsch, A., Huber, S. A., Bauer, E., Ufer, S., Plass, J. L., & Seidel, T. (2024). Effects of real-time adaptivity of scaffolding: Supporting pre-service mathematics teachers' assessment skills in simulations. *Learning and Instruction*, *94*, 101994. https://doi.org/10.1016/j.learninstruc.2024.101994

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., & Benz, C. (2021). *Didaktik der Arithmetik: Fundiert, vielseitig, praxisnah* (5., überarbeitete Auflage). Springer Spektrum.

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

Plato & Jowett, B. (1999). *Meno*. Project Gutenberg.

Radatz, H. (1980a). Students' errors in the mathematical learning process: a survey. *For the learning of mathematics*, *1*(1), 16–20.

Radatz, H. (1980b). Untersuchungen zu Fehlleistungen im Mathematikunterricht. *Journal für Mathematik-Didaktik*, *1*(4), 213–228. https://doi.org/10.1007/BF03338638

Radkowitsch, A., Sommerhoff, D., Nickl, M., Codreanu, E., Ufer, S., & Seidel, T. (2023). Exploring the diagnostic process of pre-service teachers using a simulation – A latent profile approach. *Teaching and Teacher Education*, *130*, 104172. https://doi.org/10.1016/j.tate.2023.104172

Reinhold, F., Strohmaier, A., Hoch, S., Reiss, K., Böheim, R., & Seidel, T. (2020). Process data from electronic textbooks indicate students' classroom engagement. *Learning and Individual Differences*, *83-84,* 101934. https://doi.org/10.1016/j.lindif.2020.101934

Reiss, K., & Obersteiner, A. (2019). Competence models as a basis for defining, understanding, and diagnosing students' mathematical competences. In A. Fritz, V. G. Haase, & P. Räsänen (Eds.), *International Handbook of Mathematical Learning Difficulties* (pp. 43–56). Springer International Publishing. https://doi.org/10.1007/978-3-319-97148-3_4

Reiss, K., & Winkelmann, H. (2009). Kompetenzstufenmodelle für das Fach Mathematik im Primarbereich. In D. Granzer, O. Köller, A. Bremerich-Vos, M. den van Heuvel-Panhuizen, K. Reiss, & G. Walther (Eds.), *Bildungsstandards Deutsch und Mathematik: Leistungsmessung in der Grundschule* (pp. 120–141). Beltz.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. *Educational Psychologist*, *38*(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3

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

Scherer, R., Greiff, S., & Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence*, *48*, 37–50. https://doi.org/10.1016/j.intell.2014.10.003

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, *26*(3-4), 299–323. https://doi.org/10.1080/00461520.1991.9653136

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2020). Förderung von Diagnosekompetenzen durch Scaffolding. In H.-S. Siller, W. Weigel, & J. F. Wörler (Eds.), *Beiträge zum Mathematikunterricht 2020* (p. 1535). WTM-Verlag. https://doi.org/https://doi.org/10.17877/DE290R-21544

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2022). Pre-service teachers' task selections for assessing students' misconceptions. 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* (p. 403, Vol. 4). PME.

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2023a). Teachers' diagnostic activities during task-based assessments in a digital simulation. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.), *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education* (pp. 171–178, Vol. 4). PME.

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2023b). Prozesse während der aufgabenbasierten Diagnose mathematischer Fehlvorstellungen in einer digitalen Simulation. In Goethe-

Universität Frankfurt (Ed.), *Beiträge zum Mathematikunterricht 2022* (pp. 1197–1200). WTM-Verlag. https://doi.org/10.17877/DE290R-23407

Schons, C., Obersteiner, A., Fischer, F., & Reiss, K. (2024). Toward adaptive support of pre-service teachers' assessment competencies: Log data in a digital simulation reveal engagement modes. *Learning and Instruction*, *94*, 101979. https://doi.org/10.1016/j.learninstruc.2024.101979

Schons, C., Obersteiner, A., Reinhold, F., Fischer, F., & Reiss, K. (2023). Developing a simulation to foster prospective mathematics teachers' diagnostic competencies: the effects of scaffolding. *Journal für Mathematik-Didaktik*, *44*(1), 59–82. https://doi.org/10.1007/s13138-022-00210-0

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

Sherin, M. G. (2001). Developing a professional vision of classroom events. In T. Wood, B. S. Nelson, & J. E. Warfield (Eds.), *Beyond Classical Pedagogy* (pp. 75–93). Routledge.

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

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*, 101689. https://doi.org/10.1016/j.learninstruc.2022.101689

Sommerhoff, D., Leuders, T., & Praetorius, A.-K. (2022). Forschung zum diagnostischen Denken und Handeln von Lehrkräften – Was ist der Beitrag der Mathematikdidaktik? *Journal für Mathematik-Didaktik*, *43*(1), 1–12. https://doi.org/10.1007/s13138-022-00205-x

Stadler, M., Hofer, S., & Greiff, S. (2020). First among equals: Log data indicates ability differences despite equal scores. *Computers in Human Behavior*, *111*, 106442. https://doi.org/10.1016/j.chb.2020.106442

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: a systematic literature review. *Educational Studies in Mathematics*, *104*(2), 147–200. https://doi.org/10.1007/s10649-020-09948-1

Stürmer, K., Fütterer, T., Kron, S., Sommerhoff, D., & Ufer, S. (2024). What makes a simulation-based learning environment for preservice teachers authentic? The role of individual learning characteristics and context-related features. *European Journal of Psychology of Education*, *39*(4), 3277–3299. https://doi.org/10.1007/s10212-024-00837-2

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

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

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

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

Theobald, M., Bellhäuser, H., & Imhof, M. (2018). Identifying individual differences using log-file analysis: Distributed learning as mediator between conscientiousness and exam grades. *Learning and Individual Differences*, *65*, 112–122. https://doi.org/10.1016/j.lindif.2018.05.019

Thurn, C. M., Edelsbrunner, P. A., Berkowitz, M., Deiglmayr, A., & Schalk, L. (2023). Questioning central assumptions of the ICAP framework. *npj Science of Learning*, *8*(1), 49. https://doi.org/10.1038/s41539-023-00197-4

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

Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction*, *14*(5), 445–451. https://doi.org/10.1016/j.learninstruc.2004.06.014

Vygotskiĭ, L. S., & Cole, M. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, F. H. (2021). Interpreting log data through the lens of learning design: Second-order predictors and their relations with learning outcomes in flipped classrooms. *Computers & Education*, *168*, 104209. https://doi.org/10.1016/j.compedu.2021.104209

Weimer, M. (2013). *Learner-centered teaching: five key changes to practice* (2nd ed.). John Wiley & Sons.

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

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

Wright, V. (2018). Vygotsky and a Global Perspective on Scaffolding in Learning Mathematics. In J. Zajda (Ed.), *Globalisation and Education Reforms* (pp. 123–135, Vol. 19). Springer Netherlands. https://doi.org/10.1007/978-94-024-1204-8_8

Appendix

Note:

For copyright reasons, the appendices are not included in this publication of the dissertation.