education sciences

MDPI

Article

Simulations in Teacher Education: Learning to Diagnose Cognitive Engagement

Meral Roeben [1,2,*], Johanna Vejvoda [1], Julia Murböck [1], Frank Fischer [1,3], Florian Schultz-Pernice [1,3], Anne Lohr [1], Matthias Stadler [4], Michael Sailer [5] and Nicole Heitzmann [1,3]

1 Department of Psychology, LMU Munich, 80802 Munich, Germany
2 Department for English and American Studies, LMU Munich, 80799 Munich, Germany
3 DigiLLab, LMU Munich, 80802 Munich, Germany
4 Institute of Medical Education, LMU University Hospital, LMU Munich, 80336 Munich, Germany
5 Learning Analytics and Educational Data Mining, University of Augsburg, 86135 Augsburg, Germany
* Correspondence: meral.roeben@psy.lmu.de

Abstract: Technology has shown to be beneficial for initiating cognitive engagement. In the present study, cognitive engagement was conceptualized by the ICAP framework, proposing four levels of cognitive engagement (interactive, constructive, active, passive), which can be determined from observable student activities. To initiate cognitive engagement, teachers require diagnostic skills. With this study, we aimed to foster those skills. We designed and validated a simulation with $N = 213$ pre-service teachers to investigate the validity of the simulation. Moreover, we evaluated the difficulty of diagnosing the levels of cognitive engagement within planning and implementing lessons. We used linear regressions for the validation and confusion matrices for insights into the diagnostic process. The study results show a varying difficulty of diagnosing levels of cognitive engagement due to (a) challenges in inferring the involved cognitive processes and (b) different phases of teaching. Levels of cognitive engagement that require inferential processes to identify them are more difficult to diagnose. This highlights the importance of adding scaffolds to our simulation to help pre-service teachers understand the processes of generating knowledge and co-generating knowledge. More importantly, the study reveals shortcomings of the ICAP framework and presents first suggestions for its further development.

Keywords: cognitive engagement; simulation; technology-related diagnostic skills; difficulty of diagnosing levels of cognitive engagement

check for updates

Academic Editors: Charlott Rubach and Rebecca Lazarides

Received: 25 September 2024
Revised: 7 February 2025
Accepted: 11 February 2025
Published: 20 February 2025

Citation: Roeben, M., Vejvoda, J., Murböck, J., Fischer, F., Schultz-Pernice, F., Lohr, A., Stadler, M., Sailer, M., & Heitzmann, N. (2025). Simulations in Teacher Education: Learning to Diagnose Cognitive Engagement. *Education Sciences*, *15*(3), 261. https://doi.org/10.3390/educsci15030261

1. Introduction

It has become increasingly common to use technology in lessons to support learning at school. However, the use of technology only improves the quality of a lesson if it is implemented in an effective way (Quast et al., 2021; Wekerle et al., 2022). One way to achieve such an improved lesson quality is by planning and implementing technology in a cognitively engaging way (Sailer et al., 2024; Stegmann, 2020; Wekerle et al., 2024). For this, teachers require skills to combine pedagogical knowledge (PK) with technological knowledge (TK; Koehler et al., 2013; Willermark, 2018). If teachers integrate knowledge about cognitive engagement (PK) and knowledge about the appropriate tools for their pedagogical goals (TK), they develop technological pedagogical knowledge (TPK).

Teachers use their TPK in different phases of teaching, e.g., when planning and when implementing a lesson (DCB, 2017; Ertmer & Ottenbreit-Leftwich, 2010). Ideally, when planning a lesson, teachers diagnose the levels of cognitive engagement within their lesson plans and check the alignment of the task and learning goal. When implementing a lesson, teachers diagnose their students' current cognitive engagement. Novices may struggle with planning and implementing cognitively engaging technology-supported lessons. To scaffold the acquisition of the respective skills, novices could start with identifying the cognitive engagement in existing lesson plans (Belland et al., 2017; Quintana et al., 2004). This approach may support novices in learning how to plan cognitively engaging technology-supported lessons. To train the implementation of such lessons, instead of diagnosing actual students' cognitive engagement, novices can train on representations of students' learning activities (e.g., generated students' screen-videos or products).

In the present study, the skills to diagnose cognitive engagement in technology-supported lesson plans (i.e., a lesson plan that includes digital tools or media) and in students' technology-related activities are conceptualized as *technology-related diagnostic skills*. Even though diagnostic skills are an important part of teachers' professional knowledge, they are underrepresented in pre-service teacher training (Kramer et al., 2021). As a result, in-service teachers are often not well prepared for making these diagnostic decisions in the actual professional setting (Heitzmann et al., 2019; Oser, 2001). Therefore, it may be helpful to start acquiring technology-related diagnostic skills as soon as in the first phase of teacher training (Kramer et al., 2021). Here, simulations may be a useful tool (Chernikova et al., 2020). Simulations approximate practice (Grossman et al., 2009) while, at the same time, reducing the complexity of real professional settings (Heitzmann et al., 2019). In our study, we therefore aimed at fostering technology-related diagnostic skills in a simulation-based environment. While there is evidence that simulations support the acquisition of complex skills (Chernikova et al., 2020) and that cognitive engagement enhances technology-related learning (Sailer et al., 2024), there is a research gap regarding which types of diagnostic decisions are particularly difficult to make. With this study, we want to examine the difficulty of diagnostic decisions in detail in order to find out how to efficiently and effectively design a simulation that aims at enhancing technology-related diagnostic skills regarding cognitive engagement.

1.1. Technology-Related Diagnostic Skills as Part of Technology-Related Teaching Skills

When teachers make diagnostic decisions, they usually base them on what they observe in and outside of their classroom (Schrader, 2013). A diagnosis involves classifying the forms and causes of phenomena such as cognitive engagement. These forms and causes are often latent or hidden rather than directly observable, requiring identification through visible cues (e.g., a student creating a complex mind map) and inferences based on professional knowledge. A diagnosis typically acts as a decision point that guides actions and interventions that are aimed at addressing and improving the identified problem (Heitzmann et al., 2019). Thus, diagnosing is based on inferring and categorizing observed phenomena, such as cognitive engagement. To facilitate access to information about such hidden phenomena, teachers can harness technology in various ways. In this case, technology supports teachers in terms of generating knowledge about their students. Technology can also support students. If implemented in an effective and efficient way, technology has the potential to foster students' learning processes, especially when cognitively engaging them (Sailer et al., 2024). To guarantee such a beneficial and efficient technology use, it is vital for teachers to be able to diagnose aspects like students' current levels of cognitive engagement and, if necessary, to readjust the task so that students are engaged at the desired level (DCB, 2017). The necessary diagnostic skills for this process require teachers to combine their pedagogical knowledge (PK, i.e., knowing how to cognitively engage students) and technological knowledge (TK; Koehler et al., 2013). Within the present study, we call these diagnostic activities technology-related diagnostic skills. With this term, we refer to the activity of diagnosing the cognitive engagement of learning goals, tasks, and student activities—all within a simulation. The term *technology-related* refers to a certain use of technology in terms of a digital tool or digital medium within the planning and implementing process of a lesson. However, the technology use itself is not diagnosed or analyzed. We define diagnostic skills as the ability to apply knowledge within a professional context. Thus, diagnostic skills require *conceptual* knowledge (i.e., theoretical knowledge like frameworks), as well as *action-oriented* knowledge (i.e., teachers putting theoretical knowledge into practice; Kopp et al., 2008). When planning and implementing cognitively engaging technology-supported lessons, teachers apply their action-oriented knowledge.

Planning and implementing lessons represent the first two phases of the teaching process, which are followed by the phases of evaluating and sharing lessons (DCB, 2017; Ertmer & Ottenbreit-Leftwich, 2010). We assume that each phase requires specific diagnostic skills. The first phase of teaching—planning a lesson—includes defining learning goals and tasks. Ideally, when designing cognitively engaging lesson plans, teachers reflect on the potential cognitive engagement that is required to reach a specific learning goal. Based on this learning goal, teachers create tasks that have the potential to initiate the level of cognitive engagement that is aimed at by the learning goal. Thus, putting thought into the level of cognitive engagement that matches the learning goal allows teachers to select an appropriate level of cognitive engagement for the respective tasks (Chi & Boucher, 2023). When planning lessons, teachers may also consider implementing technology (Jiménez Sierra et al., 2023). Used in a cognitively engaging way, technology is beneficial for the learning process (Sailer et al., 2024). This requires teachers to design technology-supported lessons that align with students' needs, thereby enhancing effective learning processes (Steffens, 2006; Wekerle et al., 2022). For learning goals and tasks, it is only possible to predict the cognitive engagement that may be achieved in the classroom later on. Therefore, for learning goals and tasks, only the *potential* cognitive engagement can be diagnosed. In contrast, in the implementation phase, the *actual* cognitive engagement can be diagnosed through observing students and analyzing their products. The implementation phase focuses on assessing students' activities and diagnosing whether the activities are carried out at the desired level of cognitive engagement. In this phase, the teachers monitor their students to assess different aspects of their students' behavior, skills, and cognitive processes (DCB, 2017).

We argue that the diagnostic skills that are needed for the planning and the implementation phase differ from each other. For the planning phase, teachers' diagnostic skills focus on estimating which level of cognitive engagement can be achieved by a certain learning goal and on matching it to a task. For the implementation phase, diagnostic skills in terms of observing students' activities and determining their current level of cognitive engagement are required. In the following, we will elaborate on how we conceptualize (levels of) cognitive engagement.

1.2. Cognitive Engagement

Several studies show that cognitive engagement is an important aspect of effective and efficient teaching and learning (Kunter & Voss, 2011; Sailer et al., 2024; Stegmann, 2020). More broadly, cognitive engagement describes the use of cognitive strategies (Chi et al., 2018). In this paper, we conceptualize it using the ICAP framework, which works with the acronym for four levels of cognitive engagement: interactive, constructive, active, and passive (Chi & Wylie, 2014). We decided to use the ICAP framework, as it is evidence-based (Chi et al., 2018) and offers teachers a heuristic to differentiate the complex concept of cognitive engagement. Moreover, studies found that the student activities that are proposed by the framework are beneficial for the learning process (Sailer et al., 2024; Stegmann, 2020).

According to this framework, students' learning activities can be categorized into four levels. Each level entails different cognitive processes regarding knowledge change (Wekerle et al., 2024). When carrying out a learning activity, students can be cognitively engaged on one of those four levels of engagement (i.e., passive, active, constructive, interactive). In the following, we will describe the knowledge change process and give examples of students' activities for each level of cognitive engagement.

The knowledge change process for the *passive* level is storing information. According to Chi and Wylie (2014), this information can later be recalled in a similar context. The passive level of cognitive engagement entails student activities in which students merely attend to the instructional material and absorb the information from that material. Students are, however, not overtly active, e.g., they do not take notes while listening, reading, or watching the instructional material. Typical student activities for the passive level are watching an explainer video, listening to a podcast or a lecture, and reading a text.

The *active* level of cognitive engagement, additionally to storing information, entails integrating information. Chi et al. (2018) describe that the student activities include a certain motoric action. However, not every motoric action can be determined as active engagement but only those in which students pay attention to the activity, e.g., by manipulating the instructional material. Typical active student activities are taking notes while listening to a lecture, manipulating a text by highlighting words or paragraphs, or pausing/rewinding a video.

Constructive student activities imply that the students generate knowledge that goes beyond the instructional material that is presented by the teacher (Chi & Wylie, 2014). Thus, constructive cognitive engagement goes beyond merely manipulating the instructional material; instead, students infer new output or create new products. Typical constructive activities are reflecting, creating concept maps, (self-)explaining, and taking notes or creating summaries that include their own thoughts (i.e., not just copying the text but using their own words).

Chi et al. (2018) explain that the *interactive* level of cognitive engagement entails the knowledge change processes described for the constructive level plus an interaction with a learning partner. This interaction has specific characteristics. Both partners are required to be constructively engaged, the amount of turn-taking needs to be sufficient, and the knowledge that is generated through the exchange could not have been constructed by either of the learning partners alone but emerges from the interaction. Typical student activities for the interactive level are discussions and debates or dealing with comprehension questions.

According to the framework, the levels of cognitive engagement can be determined from observing student activities. However, this may be harder for the active, constructive, and interactive level. For example, merely observing is not sufficient in order be sure whether the student activity is carried out on an active or constructive level. Teachers require additional information like the students' learning material and products. Critics also question whether the passive level oversimplifies deep learning processes, which can take place even when students are not actively doing something (Thurn et al., 2023). However, the ICAP framework is a heuristic to approximate the concept of cognitive engagement, and its assumptions are based on likelihoods (i.e., the likelihood that students are cognitively engaged on a passive level is higher if they are inactive; Chi & Boucher, 2023).

To simplify the process of determining whether students are engaged on an active, constructive, or interactive level, teachers may consider using technology. Technology stores the students' output and products and can therefore make the students' level of cognitive engagement visible (Henrie et al., 2015). For example, if the student activity is a group project, aiming at interactive engagement, the teacher can suggest that the students put their contributions down in a shared document or Etherpad (i.e., a digital notebook for collaboration). This may enable teachers to determine the students' activity more precisely.

Implementing technology has shown to be effective if these technology-supported activities are cognitively engaging (Sailer et al., 2024; Stegmann, 2020): the higher the level of cognitive engagement is, the better the effects are for the learning process (Sailer et al., 2024). Especially when teaching complex skills or contents, it is beneficial to design the (technology-supported) student activities constructively and interactively. However, the ICAP framework does not propose that every student activity be constructive and interactive; instead, the appropriate level of cognitive engagement depends on the learning goal (Chi & Boucher, 2023).

The heuristic of the ICAP framework seems straightforward. However, Chi et al. (2018) found that teachers struggle with certain levels of cognitive engagement. Differentiating the active and constructive level was found to be one challenge. Additionally, teachers struggled with designing a truly interactive learning activity. This hints towards different challenges when determining the levels of cognitive engagement and, therefore, towards a variation in difficulty for the levels. Based on the presented findings and the theoretical assumptions of the framework, we assume that it is easier to determine the passive level than the other three levels. It is only at the passive level that students' overt behavior is sufficient to diagnose the level of cognitive engagement accurately. In contrast, for the active, constructive, and interactive levels, learning output, products, or materials need to be considered, too. Thus, determining these levels requires an inferential process: the level of cognitive engagement is inferred from (a) observing the student activity and (b) analyzing additional materials. However, whether the inferential process makes these levels more difficult to determine and which other factors (e.g., the teaching phase) influence the degree of difficulty of determining the levels of cognitive engagement have not yet been systematically researched. Consequently, evidence regarding which characteristics make levels of cognitive engagement difficult needs to be generated. This evidence has the potential to inform us about the levels of cognitive engagement for which teachers need additional training and technology support.

1.3. Simulation-Based Learning Environments

Diagnosing levels of cognitive engagement is very complex. Fortunately, simulation-based learning has shown to be effective for acquiring complex skills (Cook, 2014) such as technology-related diagnostic skills. Simulations can reduce the complexity of a real teaching practice while at the same time offering a platform for practicing on authentic material (Chernikova et al., 2020). A gradual approximation of practice can be achieved by dividing practice situations into subelements and then adding more elements to increase the complexity (Grossman et al., 2009). Integrating simulations into the predominantly theoretical teacher education programs may be beneficial, as simulations allow for the gradual movement from theory into the practical situation (Chernikova et al., 2020; Heitzmann et al., 2019; Machts et al., 2024). The reduced complexity is one characteristic of simulations with only a segment of reality being depicted (Chernikova et al., 2020). Another key characteristic of simulations, suggested by Heitzmann et al. (2019), is the students' opportunity to interact with the simulation. To make sure that a simulation actually fosters the skills that it is supposed to enhance, it needs to be validated (i.e., ensuring that the simulation in the form of an assessment tool measures the skills that it is supposed to). For this, researchers gather evidence that supports a convincing case for the interpretation of the outcome data (M. Kane et al., 1999). In the present study, the assumption that participants with more prior knowledge perform better in the simulation (i.e., diagnose more accurately) than participants with lower prior knowledge is the validation criterion (Kane, 2006; Weise et al., 2020). Once simulations have successfully been validated, data from the validation study can be used to gain insights into participants' patterns within the simulation. Based on these insights, the simulation can then be turned into a learning environment.

2. The Present Study on Diagnosing Cognitive Engagement Within a Simulation

Since planning and implementing technology-supported lessons in a cognitively engaging way supports students' learning processes (Sailer et al., 2024), teachers need relevant skills to achieve this. To foster these, we created the simulation *Digivate*. With this study, we aim at validating Digivate and at gaining a deeper understanding of the ICAP framework. We validate the simulation Digivate by investigating whether higher prior knowledge is predictive of higher performance within the simulation (RQ1). We also address the lack of systematic research concerning the difficulty of diagnosing different levels of cognitive engagement (i.e., with and without inferential processes) within the planning and implementation phase (RQ2). To generate the respective evidence, we investigate which characteristics of the levels of cognitive engagement lead to challenges in the diagnostic process. Therefore, we investigate the differences in the levels of cognitive engagement (RQ2). We assume that this evidence provides us with insights into efficient ways to support acquiring technology-supported diagnostic skills regarding cognitive engagement.

RQ1. To what extent does pre-service teachers' prior knowledge predict their technology-related diagnostic skills regarding cognitive engagement within a simulation?

H1. *We hypothesize that higher prior knowledge is predictive of higher performance in technology-related diagnostic skills (i.e., a positive relation between prior conceptual knowledge and performance in the simulation).*

RQ2. To what extent does the difficulty of diagnosing levels of cognitive engagement depend on differences in the levels of cognitive engagement (inferring vs. no inferring) within the phases of teaching (i.e., planning phase, implementation phase)?

H2a. *We hypothesize that levels of cognitive engagement with no need of inferring (passive) are easier to distinguish than levels that require inferring (active, constructive, interactive; Chi et al., 2018).*

H2b. *We hypothesize that the difficulty of diagnosing the levels of cognitive engagement is different in the planning and the implementation phase* (Ertmer & Ottenbreit-Leftwich, 2010).

3. Materials and Methods

3.1. Sample and Design of the Present Study

We conducted an observational study with a correlational design as a computer-based online simulation. As there was no intervention integrated, all participants took part under the same conditions. However, the cases presented in this study were allocated to the participants in a randomized sequence. Data were collected between August 2022 and February 2023. The target group consisted of pre-service teachers, who we recruited mostly through advertising in courses at university. Participants were also given the opportunity to receive monetary compensation for completing the online study. Our sample was non-representative. The total sample included N = 274 pre-service teachers. We excluded participants who did not complete the whole study and reached a final sample of n = 213 pre-service teachers. In this final sample, 76% of the participants stated that they were female, 22% male, 1% diverse, and 1% did not specify their gender. On average, the participants reported a teaching experience of almost five months (M = 4.69; SD = 6.27). While there are certain obligations regarding internships in the participants' study programs in the federal state of Bavaria, the number of lessons that are actually taught varies for every individual, with high flexibility as to how the teaching experience is gained. Therefore, we cannot make a statement as to where the reported teaching experience originates from. In total, 45.5% of the participants stated that they were unfamiliar with the ICAP framework, and 84% reported to never have used the ICAP framework before.

3.2. Learning Environment and Learner Task

For our study, we developed the simulation-based learning environment *Digivate*, which we designed as a point-and-click adventure. The setting of the simulation is a secondary school. At this school, teacher trainees complete their apprenticeship to become teachers. The school is currently dedicating one week to the topic of sustainability. The fictive seminar teacher, Klara Sinn, introduces the teacher trainees to this setting. She explains that they will first look at existing lesson plans and analyze the potential level of cognitive engagement of learning goals and tasks (planning phase) and then visit a class (implementation phase). In the classroom, the teacher trainees (i.e., study participants) will assess the students' level of cognitive engagement while the students are working on their tasks. Before diagnosing the levels of cognitive engagement, participants are offered to watch an introductory video on the ICAP framework. Watching the video was voluntary, as we know the ICAP framework to be part of one course within the teacher education program. With our participants studying in various semesters, we assumed differences in prior knowledge regarding the framework. To avoid redundancy, we decided on a voluntary input but urged participants to watch the video if they had not taken part in said course.

The simulation is divided into two main parts: the planning phase (see Figure 1) and the implementation phase (see Figure 2). In both phases, participants work on cases. Cases in the simulation consist of either diagnosing one lesson plan or one student activity. Cases in both the planning and the implementation phase were randomly picked and allocated to the participants out of a total of 15 cases. In the *planning phase*, a case consists of a lesson plan (i.e., learning goal, task, social grouping, and media) for which the potential level of cognitive engagement of the learning goal and the task are to be diagnosed (see Figure 1). A case in the *implementation phase* is designed as follows: First, a picture of the classroom with students sitting at desks is shown. One of the students has a colored square around their head. Clicking on this square directs the participant to the screen-video depicting what this student is currently doing on their screen (see Figure 2). The product that results from the activity in the video is shown on the next slide. Based on this input, participants choose at which level of cognitive engagement the student is currently engaged. In the sidebar, the screen-video, the learning product, the corresponding lesson plan with its desired levels of cognitive engagement for the learning goals, and the ICAP framework slides can be accessed for further support.

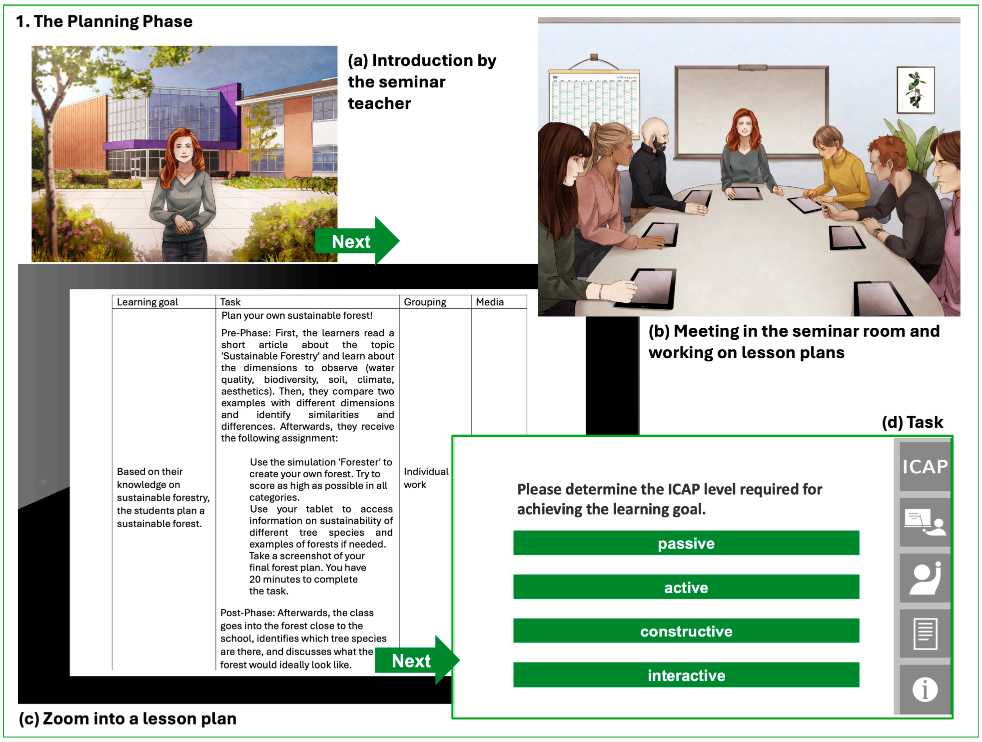

Figure 1. Overview of the structure of the planning phase.

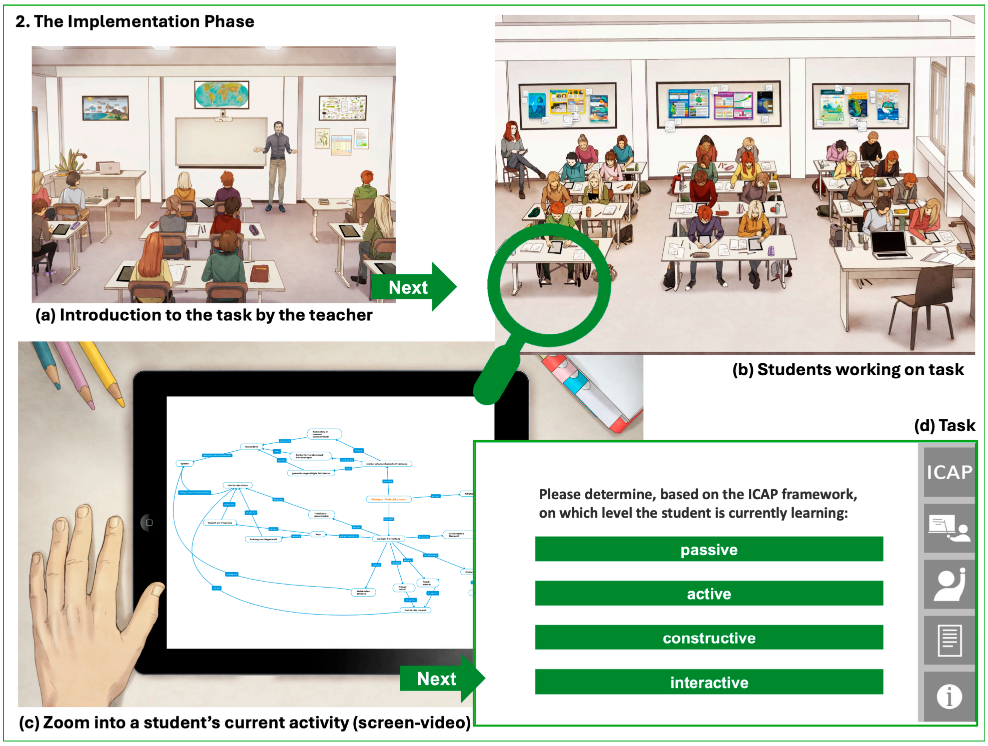

Figure 2. Overview of the structure of the implementation phase.

After completing the first case of the planning phase, the participants were assessed on their prior knowledge on cognitive engagement. They were not provided with any elements for assistance during the test. Having completed the test, the participants worked on the remaining six cases of the planning phase. After completing all cases of the planning phase, the participants found themselves back with the seminar teacher, who showed them the solution to all the lesson plans. In the next step, the seminar teacher informed the participants that they would now see how the lesson plans were implemented in a classroom. For this implementation phase, the participants found themselves in a simulated classroom with a teacher and students (see Figure 2). The teacher introduced the teacher trainees to the class and then showed their students their next task. The study participants were instructed to walk around and determine the students' level of cognitive engagement. In total, participants worked on six cases in the implementation phase.

3.3. Measures Within the Present Study

We assessed prior knowledge with a single-choice test on conceptual knowledge on cognitive engagement. Within the simulation, we determined technology-related diagnostic skills by assessing the diagnostic accuracy.

(1) Prior Knowledge

We assessed the prior knowledge with eight single-choice items on domain-specific (Süß & Kretzschmar, 2018) *conceptual* knowledge of cognitive engagement that we designed ourselves. The test shows acceptable reliability (McDonald's Omega ω = .69; Stadler et al., 2021). Each item offered four possible answers. The test included items like "In which of the four learning activities according to the ICAP framework, no product is produced?".

(2) Technology-Related Diagnostic Skills (Diagnostic Accuracy)

To measure technology-related diagnostic skills, we assessed the participants' performance in the simulation. The performance was measured by diagnostic accuracy (i.e., the sum of correctly identified levels of cognitive engagement in single-choice items; Braun et al., 2017; Hege et al., 2018). Items included the following tasks: "Please assess which ICAP level is necessary for achieving the learning goal."; "Please assess which ICAP level is necessary for achieving the learning task."; and "Based on the ICAP framework, please assess, on which ICAP level the student is learning right now". For each single-choice item, participants could achieve zero or one points. Thus, a maximum of two points per case could be gained in the planning phase and a maximum of one point in the implementation phase. In contrast to the *conceptual* knowledge that was assessed by the single-choice test (i.e., prior knowledge), we assessed the technology-related diagnostic skills through participants *applying* their knowledge (i.e., *action-oriented* knowledge) on cognitive engagement (Kopp et al., 2008). Thus, participants did not only need to know the levels of cognitive engagement and their characteristics but had to apply and transfer this knowledge in order to diagnose the simulated cases.

(3) Levels of Cognitive Engagement

The levels of cognitive engagement were represented by the levels of the ICAP framework (passive, active, constructive, interactive). The accurate levels of cognitive engagement for each learning goal, task, and student activity were validated in an expert workshop. We differentiated between levels of cognitive engagement that require inferential processes (active, constructive, interactive) and levels of cognitive engagement that do not require inferring (passive). We also differentiated between the levels of cognitive engagement in the planning phase and the levels of cognitive engagement in the implementation phase.

(4) Difficulty of Diagnosing Levels of Cognitive Engagement

We operationalized the difficulty of diagnosing cognitive engagement with sensitivity and specificity. Sensitivity and specificity can be calculated with the help of confusion matrices. Confusion matrices compare the accurate answer (predicted level) with the selected answer of participants (actual level). Sensitivity and specificity indicate which levels of cognitive engagement are more challenging to determine. Sensitivity is the proportion of true positive values to all other answers for this predicted level. Sensitivity describes how effectively a test correctly identifies the correct answer. Specificity is the proportion of true negative values to the sum of true negative and false positive values. Specificity describes how effectively a test identifies the inaccurate answers as inaccurate.

3.4. Statistical Analyses to Address the Research Questions of the Present Study

To address RQ1, we conducted linear regression analyses, investigating whether higher prior knowledge is predictive of higher technology-related skills.

To address RQ2, we used confusion matrices to measure the difficulty of diagnosing cognitive engagement. Confusion matrices are matrices comparing predicted values (columns) with the actual values (rows). If participants correctly chose (actual level) the accurate level of cognitive engagement (predicted level), this was a true positive value (TP). If they chose an inaccurate level of cognitive engagement (actual level), this was a false negative value (FN). All levels of cognitive engagement that were correctly not chosen were true negative values (TNs). False positive values encompassed the wrongly selected predicted levels of cognitive engagement. For example, if active was the predicted level, if a participant chose active (actual value) although the predicted level was passive, constructive, or interactive, this level was called a false positive (FP).

We used chi-square tests to determine whether the confusion between levels of cognitive engagement was significant. For testing whether there were significant differences between sensitivities and specificities, we used confidence intervals (Wilson method). If there was no overlap of the intervals, there was a significant difference between the values.

4. Results

4.1. Prior Knowledge and Performance (RQ1)

The descriptive results (see Table 1) show that the participants have moderate to good prior knowledge and that their diagnostic accuracy (i.e., their performance in the simulation) was also moderate to good.

Table 1. Descriptive results for prior knowledge and diagnostic accuracy.

Variables	*N*	*M*	*SD*	*Min*	*Max*
Prior Knowledge	213	5.27	2.04	0	8
Diagnostic Accuracy	213	10.05	3.25	3	16

Linear regressions show that participants with higher prior knowledge diagnosed more accurately, with prior knowledge accounting for 21% of the variance in diagnostic accuracy (R^2 = .21). The standardized regression coefficient for prior knowledge of β = .45 (*S.E.* = .10; p < .001) indicates a moderately strong and statistically significant positive effect on the performance in the simulation. These findings support the validity claim with respect to RQ1.

4.2. Varying Difficulty Depending on the Levels of Cognitive Engagement and Phases of Teaching (RQ2)

To address RQ2, we created confusion matrices for the simulation over both phases (planning and implementation phase) and separate confusion matrices for the planning phase and the implementation phase.

Table 2 depicts the confusion matrix for the simulation overall, comparing the predicted (accurate) levels with the actual (selected by participants) levels. We will now describe each level of cognitive engagement within the confusion matrix.

Table 2. Confusion matrices for both phases of teaching: predicted vs. selected levels of cognitive engagement.

		Predicted			
		P	A	C	I
Actual	P	224 (63%)	172 (15%)	93 (7%)	6 (1%)
	A	65 (18%)	541 (48%)	316 (24%)	73 (7%)
	C	39 (11%)	253 (22%)	693 (52%)	140 (14%)
	I	28 (8%)	159 (14%)	240 (18%)	790 (78%)

The *passive* level was twice as often confused with active as with interactive. *Active* was mostly confused with constructive, followed by passive and, to a lesser extent, interactive. The *constructive* level was diagnosed accurately in more than half of the cases. The constructive level was most frequently confused with the active level, followed by the interactive and, lastly, the passive level. In the majority of cases, the *interactive* level was accurately determined, with its true positives standing out in contrast to the other levels of cognitive engagement. The interactive level was sometimes confused with the constructive level.

To sum up, Table 2 shows that active and constructive are confused with each other more frequently than any other levels are confused with each other. The 4 × 4 chi-square tests confirm a significant association between the active and constructive levels ($\chi 2$ (1) = 240.05, $p < .001$) and a significant association between the constructive and interactive levels ($\chi 2$ (1) = 658.41, $p < .001$). This indicates that levels of cognitive engagement that need to be inferred (active, constructive, interactive) are more difficult to distinguish from each other than the directly observable passive level.

Taking a look at sensitivity and specificity, the interactive level shows the highest sensitivity (see Table 3), meaning that it is very likely that the accurate level of cognitive engagement (interactive) is selected by the study participants. The second highest sensitivity can be found for passive, followed by constructive and active. Thus, for active, the likelihood that the accurate level (active) is selected correctly is relatively low. The confidence intervals (Wilson method) show that only for active (.45–.51) and constructive (.49–.54) there is no significant difference in sensitivity, as their confidence intervals overlap. This indicates that these two levels (active and constructive) are especially and similarly hard to diagnose, which partly supports our hypothesis: levels of cognitive engagement that need to be inferred are more difficult to determine. However, interactive also needs to be inferred but shows a high sensitivity.

Table 3. Sensitivity and specificity for the simulation over both phases of teaching.

	Sensitivity	**Specificity**
Passive	63%	92%
Active	48%	83%
Constructive	52%	83%
Interactive	78%	85%

Sensitivity describes how likely it is that participants correctly select the accurate level of cognitive engagement. However, not only does sensitivity provide us with information about the different difficulties of determining different levels of cognitive engagement, but so does specificity. Specificity is high for all four levels (see Table 3), indicating that it is highly likely that a level is determined as inaccurate when it is in fact inaccurate. A significant difference in specificity (Wilson method) can be found between passive (.91–.93) and the other three levels: active (.82–.85), constructive (.81–.84), and interactive (.84–.86).

To test hypothesis H2b, which predicts that the planning and implementation phases differ in difficulty when diagnosing levels of cognitive engagement, we looked at separate confusion matrices for the planning and the implementation phase (see Tables 4–6). We created two confusion matrices for the planning phase: one for the learning goals (see Table 4) and one for the tasks (see Table 5).

Table 4. Confusion matrix for the planning phase (learning goals): predicted vs. selected levels of cognitive engagement.

		Predicted			
		P	A	C	I
Actual	P	NA	88 (21%)	85 (14%)	1 (0%)
	A	NA	180 (44%)	172 (29%)	25 (9%)
	C	NA	109 (26%)	260 (44%)	47 (17%)
	I	NA	35 (8%)	73 (12%)	203 (74%)

Table 5. Confusion matrix for the planning phase (tasks): predicted vs. selected levels of cognitive engagement.

		Predicted			
		P	A	C	I
Actual	P	153 (59%)	50 (11%)	5 (1%)	0 (0%)
	A	46 (18%)	274 (58%)	110 (23%)	20 (7%)
	C	33 (13%)	81 (17%)	289 (60%)	42 (15%)
	I	27 (10%)	66 (14%)	79 (16%)	214 (78%)

Within the *planning phase*, participants started by diagnosing the level of cognitive engagement of the *learning goal*. None of the lesson plans included learning goals that were aimed at a passive level of cognitive engagement. As Table 4 shows, *active* was most frequently confused with constructive, followed by passive and interactive. The *constructive* level was mostly confused with the active level. *Interactive* was confused most frequently with constructive, followed by active and passive.

Table 5 depicts the confusion matrix for *tasks* (i.e., comparing the predicted levels of cognitive engagement of tasks to the selected levels by participants). The *passive* level of cognitive engagement was mostly determined accurately. The *active* level was slightly more often confused with constructive than with interactive and passive. The *constructive* level was mostly confused with the active level. *Interactive* was mostly accurately diagnosed and sometimes confused with the constructive level.

To summarize, Tables 4 and 5 demonstrate that for the planning phase (learning goals and tasks), the constructive and active levels, which are both levels that need to be inferred, are most frequently confused with each other. This result was confirmed in 4×4 chi-square tests (learning goals: $\chi 2$ (1) = 34.09; $p < .001$; tasks: $\chi 2$ (1) = 183.05; $p < .001$), which partly supports H2a, as the active and constructive levels need to be inferred, but so does interactive, which is not confused as often in the planning phase.

Looking at the sensitivity for the planning phase (see Table 7), we found different results for learning goals and tasks. Learning goals showed a rather low sensitivity for active and constructive but a higher one for interactive. This indicates that participants struggled with correctly diagnosing the accurate levels of learning goals, in particular the active and constructive levels. In contrast to learning goals, tasks overall showed a higher sensitivity. The sensitivity for tasks was highest for interactive, followed by constructive, passive, and active. To evaluate whether the sensitivities significantly differed, we used the Wilson method to generate confidence intervals. In the planning phase (learning goals and tasks), the confidence intervals for the sensitivities only showed significant differences in sensitivity between interactive (.68–.78; .72–.82) and the other three levels of cognitive engagement: passive (NA[1]; .53–.65), active (.39–.49; .54–.63), and constructive (.40–.48; .55–.64). The confidence intervals for passive, active, and constructive overlapped, indicating no significant difference in their sensitivity within the planning phase. This suggests that the interactive level differs significantly in its difficulty compared to the other levels. Since interactive showed the highest sensitivity for learning goals and tasks, we can assume that it is significantly easier to determine an interactive level, which contradicts our assumption that all levels of cognitive engagement that need inferring are more difficult to determine.

For learning goals, the specificity of the interactive level was the highest, followed by passive, active and constructive. This indicates that it was likely that participants did, in fact, identify (select) the interactive level as inaccurate. The confidence intervals (Wilson method) for learning goals showed no significant difference in specificity between active (.74–.80) and constructive (.74–.80) and for passive (.84–.88) and interactive (.87–.91), as the confidence intervals overlapped. The confidence interval of interactive (.87–.91) did not overlap. For tasks, passive showed the highest specificity, followed by interactive, constructive, and active. For tasks, only the confidence interval for passive (.94–.97) did not overlap with the other levels (active: .80–.85; constructive: .82–.87; interactive: .84–.88), indicating a significant difference in specificity between passive and the other levels. This supports H2a, as this result proposes a lower difficulty for the passive level, a level that does not require inferential processes.

To sum up, the sensitivities of active and constructive are below the sensitivity values of the interactive level (see Table 7), and the significant results in chi-square tests indicate that within the planning phase, the active and constructive levels of cognitive engagement are most difficult to differentiate. This partly supports our H2a, as the active and constructive levels require inferential processes and are more difficult to determine. However, the results indicate that interactive, also a level that needs inferring, is rather easy to determine within the planning phase.

Table 6. Confusion matrix for the implementation phase: predicted vs. selected levels of cognitive engagement.

		Predicted			
		P	A	C	I
Actual	P	71 (73%)	34 (14%)	3 (1%)	5 (1%)
	A	19 (20%)	87 (36%)	34 (13%)	28 (6%)
	C	6 (6%)	63 (26%)	144 (54%)	51 (11%)
	I	1 (1%)	58 (24%)	88 (3%)	373 (82%)

In a third step, we looked at the confusion matrix for the *implementation phase* (see Table 6), depicting the accurate (predicted) and selected (actual) levels of cognitive engagement for student activities. The *passive* level was usually accurately diagnosed. *Active* was most often confused with constructive, closely followed by interactive, but only a few times with passive. *Constructive* was mostly confused with interactive and less so with active. The *interactive* level for tasks was rarely confused with other levels. In contrast to the planning phase, in the implementation phase, the most frequently confused levels of cognitive engagement were constructive and interactive. This was confirmed by a 4 × 4 chi-square test ($\chi 2$ (1) = 177.36, p < .001). The sensitivities in the implementation phase were highest for the passive and interactive levels (see Table 7), indicating that these levels of cognitive engagement were easier to diagnose accurately than the active and constructive levels. The confidence intervals (Wilson method) overlapped for the sensitivity of passive (.64–.81) and interactive (.78–.85). This indicates significant differences in difficulty for diagnosing the easier levels of passive and interactive compared to the more difficult levels of active (.30–.42) and constructive (.48–.59). Regarding specificity, that of passive was the highest; interactive showed the lowest specificity. In student activities, the differences in specificity were significant between all levels of cognitive engagement, as there was no overlap in their confidence intervals. The low specificity of interactive in contrast to its high sensitivity indicates that it is easy to determine the interactive level when students act interactively, but it is difficult to recognize that students are not engaged on an interactive level. This can also be observed in the confusion matrix for the implementation phase (see Table 6) when looking at the column for constructive. Constructive was confused with interactive 88 times, which means that student activities that represented cognitive engagement on a constructive level were inaccurately diagnosed as interactive 88 times. Cognitive engagement on a constructive level involves generating knowledge on one's own, while cognitive engagement on an interactive level involves generating knowledge with others. Thus, the low specificity of the interactive level and the frequent confusing of constructive with interactive might lead to the conclusion that participants diagnosed an interactive level because more than one person was involved in the activity although no knowledge was co-generated. The findings concerning the interactive level in the implementation phase suggest that there is another difference in difficulty between the two phases of teaching.

Table 7. Sensitivity and specificity for planning and implementation phase.

	Planning Phase				Implementation Phase	
	Learning Goal		Task		Student Activity	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Passive	NA	86%	59%	96%	73%	96%
Active	44%	77%	58%	83%	36%	90%
Constructive	44%	77%	60%	84%	54%	85%
Interactive	74%	89%	78%	86%	82%	76%

5. Discussion

5.1. Summary of the Results

With the present study, we aimed at gaining insights regarding ways to support pre-service teachers in acquiring technology-related diagnostic skills. For this, we validated the simulation Digivate (RQ1). Moreover, we systematically investigated the difficulty of levels of cognitive engagement that require inferential processes and those that do not require them within two phases of teaching (i.e., the planning phase and implementation phase; RQ2).

With respect to RQ1, we found that prior knowledge is predictive of performance in technology-related diagnostic skills. This result is in support of H1 and, therefore, our validity claim. With respect to RQ2 (i.e., investigating the difficulty of assessing the different levels of cognitive engagement), we found that active and constructive levels of cognitive engagement, as well as constructive and interactive levels, are difficult to distinguish from each other and therefore are difficult to determine. This supports H2a, which posits that levels of cognitive engagement that do not need to be inferred are easier to determine than those that need to be inferred. Moreover, H2b (i.e., different diagnostic difficulties in the phases of teaching) is also supported, as the confusion of active and constructive occurred in the planning and the confusion of constructive and interactive in the implementation phase. Another indication is that for learning goals (planning phase), it was not only difficult to determine the accurate active and constructive level of cognitive engagement as being accurate. For learning goals, it was also difficult to identify inaccurate active and constructive levels as being inaccurate.

5.2. Practical and Theoretical Implications of the Results

From the results, we can derive implications (a) for supporting the acquisition of technology-related diagnostic skills and (b) for the design of teacher education programs at universities. For (a), based on finding the simulation to be valid and further findings about its specific difficulties for participants, we can draw conclusions towards the way in which the acquisition of technology-related diagnostic skills could be supported more effectively and, thus, how we could redesign our simulation and turn it into a learning environment. Our findings concerning the frequent confusion of the active and constructive levels align with what Chi et al. (2018) found: instead of inspecting the content of students' products, teachers took the product itself as an indicator of knowledge generation and therefore diagnosed a constructive level. Whether knowledge was generated is, however, the main difference between active and constructive. It is important to know that pre-service teachers struggle with diagnosing knowledge generation to be able to support their understanding of the knowledge generation process. This lack of understanding could, for example, be tackled by incorporating elaborate feedback on their diagnostic decisions. In the implementation phase, participants tended to diagnose that students were interactively engaged when they were in fact constructively engaged. This finding shows that it was difficult for participants to recognize an inaccurate interactive level as being inaccurate but not to determine an accurate interactive level correctly as accurate. Cognitive engagement on an interactive level implies that knowledge is co-generated through students combining their contributions. However, it may be difficult to decide in a conversation if every student generates knowledge on their own (constructive) or through co-construction (interactive). Determining whether knowledge is generated alone or together might be easier in lesson plans. Lesson plans include learning goals and tasks, and both usually describe the intended activity in detail. In contrast, screen-videos of students' activities may only depict this important aspect of the interactive level implicitly. For example, in the implementation phase, several students may talk to each other about the design of a sustainable forest, and new knowledge is generated. However, if each student comes up with their own combination of trees, the knowledge is not generated in the process, but each student generates knowledge independently and shares it with the other students. Participants working on the simulation in the present study seemed to overgeneralize the obvious feature of more than one person being involved in the implementation phase, as they frequently confused the constructive and interactive levels. Simulations offer valuable features to confront this overgeneralization of the more-than-one-person-being-involved characteristic as they bear the possibility to manipulate details of simulated cases. For example, one simulated case could represent an activity that involves several people. While this can be kept constant, the way in which they interact can be repeatedly manipulated. This way, accurately distinguishing constructive from interactive levels of cognitive engagement can be explicitly trained, as well as facilitated. This way, covert cognitive processes such as knowledge (co-)generation can be made salient within a simulation. This enables pre-service teachers to practice inferring these processes and increases the likelihood of being able to apply these inferential skills when finally standing in a classroom.

When analyzing differences between the difficulty in diagnosing levels of cognitive engagement between the planning and the implementation phase (RQ2), we found different results for the two components (i.e., learning goals and tasks) of the planning phase. For learning goals, it was more difficult to determine the accurate level of cognitive engagement as accurate (i.e., low sensitivity) than it was for tasks. In the implementation phase, similarly to tasks in the planning phase, it was easier to determine accurate levels as accurate (i.e., high sensitivity) compared to determining learning goals. This hints towards challenges when combining learning goals and the ICAP framework. The ICAP framework assumes that the level of cognitive engagement can be determined either from merely observing a student's overt behavior or from additionally taking learning products into account. The screen-videos in the implementation phase align closely with the descriptions of the levels of cognitive engagement that were proposed by Chi et al. (2018), and so do tasks to a certain extent. Tasks describe the activity in detail and report on the products that are supposed to emerge. In contrast, learning goals describe an indicator that a certain skill has been acquired. Therefore, the inferential process necessary to determine a learning goal seems more demanding, as neither an activity nor a product can be directly observed or assumed. One idea to reduce this high level of difficulty might be to suggest certain verbs that are typically used for learning goals and categorize them according to the levels of cognitive engagement, similarly to the directive verbs that Chi and Boucher (2023) proposed. Another aspect that might cause challenges with learning goals but also overall is the simplicity of the ICAP framework. With its clearly separated four levels, the ICAP framework works with categories and leaves little room for a more fine-grained differentiation. This would be possible if the levels were seen as dimensions instead of categories (Wekerle et al., 2024). Some learning goals, tasks, or student activities may be more centered within the dimension of one level than others. Possibly, typical learning goals, tasks, or student activities which are in the center of a certain level may be easier to diagnose than those that are close to the border of other levels. Thus, a suggestion to advance the ICAP framework would be to add more fine-grained levels of cognitive engagement (Wekerle et al., 2024).

In contrast to learning goals, the screen-videos representing students' activities depicted the levels of cognitive engagement more directly, as technology was used to make cognitive engagement visible to the teacher (i.e., study participant). For example, they may have observed a group of students working on a task using Etherpad (a collaborative synchronous writing tool), generating knowledge during this process. Through this observation, the advantage of the tool Etherpad for fostering cognitive engagement and, therefore, for the learning process becomes visible to the participants. The study participants would not be able to identify the level of cognitive engagement as easily without the Etherpad, making the cognitive processes visible. In contrast to complex analog group work, technology can depict relevant aspects like turn-taking or the contributing author in a shared document. To be able to diagnose the cognitive engagement of students based on the screen-videos, participants need to combine their PK (diagnosing cognitive engagement) and TK (observing students who are using technology). In the simulation, we showed how technology can be used to support cognitive engagement among students (Wekerle et al., 2024) and how teachers can use technology to diagnose their students' level of cognitive engagement. This way, we aimed to foster the study participants' TPK and, therefore, to foster their technology-related teaching skills.

For (b), implications towards the design of teacher education programs at university, we suggest that more focus and awareness should be put towards how high-level cognitive engagement can be achieved. For example, it seems to be a challenge for pre-service teachers to understand what a knowledge generation process is. Therefore, we should offer more practice opportunities to pre-service teachers, enabling them to move beyond teaching at the active and passive levels (Sailer et al., 2024). There is also a need for sensitizing pre-service teachers towards the challenges of designing group work. By referring to the interactive level, clear criteria of what makes these types of exchanges beneficial for learning (i.e., every student is constructively engaged, sufficient amount of turn-taking) can be presented and applied (Chi & Wylie, 2014). However, it is also an opportunity to show that not every student is supposed to reach the interactive level all the time when collaborating. To acquire those skills, methods that are close to professional practice, like simulations, have significant potential (Chernikova et al., 2020). Teaching methods like simulations may also reduce the amount of passive learning activities in teacher education and support formative assessment (Black & Wiliam, 2009).

5.3. Limitations Within the Present Study

We assessed the prior knowledge using a pre-test on conceptual knowledge of cognitive engagement (i.e., the ICAP framework). The merely moderate reliability of this test may suggest considering our results with caution. However, we argue that cognitive engagement is a latent construct. As items of latent constructs are supposed to depict various aspects of the construct, a moderate reliability is justified (Stadler et al., 2021).

Another limitation is that, so far, we only included pre-service teachers as participants. Future studies may include even more heterogeneous groups of participants (e.g., in-service teachers vs. pre-service teachers) to validate the scope of participants who are likely to benefit from learning by using the simulation.

As during our study, COVID-19 restrictions were still in place, we were not able to conduct a laboratory study but merely a standardized online study. This reduced the possibilities for controlling for additional variables. Although the statistics indicate that our data are sufficient to be used to address the research questions, the study may be replicated under more controlled conditions. This way, we can ensure the quality of the data more proficiently, since we assume that participants are less likely to be interrupted or pause when working on the simulation in laboratory conditions.

Moreover, although interaction of the study participants with the simulation is a key characteristic of simulations (Heitzmann et al., 2019), for the validation, the interaction possibilities of the simulation were significantly restricted. Participants were guided through the simulation on a predetermined pathway without the possibility to repeat and practice or further explore. This decision resulted from our validation criterion, as we wanted to investigate whether prior knowledge is predictive of the performance in the simulation. In a future study, it would be interesting to explore whether comparable results occur with more degrees of freedom for the participants to interact and choose different paths within the simulation.

The present study contributed to advancing the ICAP framework. The proposed changes, however, indicate limitations in the framework. First, the ICAP framework is very focused on the implementation phase of teaching, making it challenging to evaluate learning goals based on its proposed levels of cognitive engagement. In addition, each level of cognitive engagement is rather wide (Wekerle et al., 2024). More fine-grained levels may reduce the diagnostic difficulty, make learning goals more feasible within the ICAP framework, and reduce the proposed hierarchy of the levels. Lastly, although the framework suggests that levels of cognitive engagement can be determined based on the observable student activities, this is challenging when it comes to the active, constructive, and interactive levels. These levels involve covert cognitive processes requiring inferring. To practice this, simulation-based learning seems to be an effective and efficient extension to teacher education programs.

6. Conclusions

We conclude that the simulation is sufficiently valid to be used as a learning environment for pre-service teachers in higher education. We also conclude that models of cognitive engagement in technology-supported lessons (e.g., the ICAP framework; Chi & Wylie, 2014) are, in principle, relevant and helpful for supporting pre-service teachers in learning to plan and implement technology-supported instruction for their classroom. However, these models need to be revised as well.

Findings like challenges regarding the process of (co-)generating knowledge, diagnosing learning goals, or the lack of fine-grained levels (Wekerle et al., 2024) provide insights regarding the design of an updated, adaptive version of the simulation (Plass & Pawar, 2020). Regarding the difficulty of inferring levels of cognitive engagement, scaffolds could support pre-service teachers (Belland et al., 2017). This could be achieved by making the covert learning processes more salient (Machts et al., 2024). For example, the knowledge generation can be made more visible by highlighting important aspects in the screen-videos or by adding written descriptions that make the processes more explicit. To practice the skills of accurately diagnosing the interactive level, similar cases, each involving more than one student, that systematically discriminate the level of cognitive engagement between constructive and interactive can be used. Additionally, support for diagnosing learning goals is needed. An updated simulation may also include more degrees of freedom (Heitzmann et al., 2019) and possibilities for self-regulated learning (Bannert, 2009). Such an updated version of Digivate can be implemented in teacher education as a formative assessment. This training opportunity may foster pre-service teachers' skills and awareness regarding the benefits that come with cognitively engaging their students, which will hopefully result in them planning and carrying out lessons beyond the passive and active levels of cognitive engagement.

However, not only pre-service teachers benefit from such a learning environment, but also in-service teachers. Thus, implementing the learning environment is important in all three phases of teacher education (i.e., in the first phase at university, in the second phase during practical teacher training, and for in-service teacher trainings).

Author Contributions: Conceptualization, M.R., J.V., J.M., F.F., F.S.-P., A.L., M.S. (Matthias Stadler), M.S. (Michael Sailer) and N.H.; methodology, M.R., J.M., F.F., M.S. (Matthias Stadler), M.S. (Michael Sailer) and N.H.; validation, M.R., J.V., J.M., F.F., F.S.-P., A.L., M.S. (Matthias Stadler), M.S. (Michael Sailer) and N.H.; formal analysis, M.R., J.M., M.S. (Matthias Stadler) and N.H.; investigation, M.R., J.V., J.M., F.F., F.S.-P., A.L., M.S. (Matthias Stadler), M.S. (Michael Sailer) and N.H.; data curation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, M.R., J.V., J.M., F.F., F.S.-P., A.L., M.S. (Matthias Stadler), M.S. (Michael Sailer) and N.H.; visualization, M.R.; supervision, F.F. and N.H.; project administration, J.M., F.F., F.S.-P., M.S. (Michael Sailer) and N.H.; funding acquisition, F.F. All authors have read and agreed to the published version of the manuscript.

Funding: The first author's contribution by Meral Roeben was funded by a grant from the Hanns Seidel Foundation. Further contributions to this research by Anne Lohr and Julia Murböck were funded by a grant from the German Federal Ministry of Research and Education under the grant number 01JA1810. Contributions to this research by Frank Fischer, Florian Schultz-Pernice, and Johanna Vejvoda was funded by the European Union—NextGenerationEU—and by the German Federal Ministry of Education and Research under the grant number 01JA23S01E. The authors are responsible for the content of this publication.

Institutional Review Board Statement: An ethical review was not required for the present study, as it was purely observational and did not involve any experimental intervention or manipulation. Participation was entirely voluntary, and no risks, harm, or disadvantages were associated with the study. No sensitive personal data were collected, and any identifying information was stored separately from the study data to ensure anonymity.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author, as we plan on further studies based on this data set.

Acknowledgments: The authors thank Julia Engler for her support in designing the simulation and the first data analyses. They also thank Thomas Weber and Anton Mai for their support in the technological implementation of Digivate.

Conflicts of Interest: The authors declare no conflicts of interest.

Note

[1] The lesson plans did not include learning goals that aimed at generating a potential cognitive engagement at the passive level.

References

Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie*, *23*(2), 139–145. [CrossRef]

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. [CrossRef] [PubMed]

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*, 5–31. [CrossRef]

Braun, L. T., Zottmann, J. M., Adolf, C., Lottspeich, C., Then, C., Wirth, S., Fischer, M. R., & Schmidmaier, R. (2017). Representation scaffolds improve diagnostic efficiency in medical students. *Medical Education*, *51*(11), 1118–1126. [CrossRef]

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. [CrossRef]

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. [CrossRef]

Chi, M. T. H., & Boucher, N. S. (2023). Applying the ICAP framework to improve classroom learning. In C. E. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In their own words: What scholars and teachers want you to know about why and how to apply the science of learning in your academic setting* (pp. 94–110). Society for the Teaching of Psychology.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. [CrossRef]

Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education*, *48*(8), 750–760. [CrossRef]

DCB (Forschungsgruppe Lehrerbildung Digitaler Campus Bayern) [Research Group Teacher Education Digital Campus Bavaria]. (2017). Kernkompetenzen von lehrkräften für das unterrichten in einer digitalisierten welt [Core competencies of teachers for teaching in a digital world]. *Merz Medien+ Erziehung: Zeitschrift für Medienpädagogik*, *4*, 65–74. [CrossRef]

Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change. *Journal of Research on Technology in Education*, *42*(3), 255–284. [CrossRef]

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

Hege, I., Kononowicz, A. A., Kiesewetter, J., & Foster-Johnson, L. (2018). Uncovering the relation between clinical reasoning and diagnostic accuracy—An analysis of learner's clinical reasoning processes in virtual patients. *PLoS ONE*, *13*(10), e0204900. [CrossRef]

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 in higher education: A framework and a research agenda. *Frontline Learning Research*, *7*(4), 1–24. [CrossRef]

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. [CrossRef]

Jiménez Sierra, Á. A., Ortega Iglesias, J. M., Cabero-Almenara, J., & Palacios-Rodríguez, A. (2023). Development of the teacher's technological pedagogical content knowledge (TPACK) from the lesson study: A systematic review. *Frontiers in Education*, *8*, 1078913. [CrossRef]

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, *18*(2), 5–17. [CrossRef]

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Praeger.

Koehler, M. J., Mishra, P., & Cain, W. (2013). What is technological pedagogical content knowledge (TPACK)? *The Journal of Education*, *193*(3), 13–19. [CrossRef]

Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, case-based worked examples: Effects of erroneous examples and feedback. *Medical Education*, *42*(8), 823–829. [CrossRef] [PubMed]

Kramer, M., Förtsch, C., Seidel, T., & Neuhaus, B. J. (2021). Comparing two constructs for describing and analyzing teachers' diagnostic processes. *Studies in Educational Evaluation*, *68*, 100973. [CrossRef]

Kunter, M., & Voss, T. (2011). Das modell der unterrichtsqualität in COACTIV: Eine multikriteriale analyse [The framework of teaching quality in COACTIV: A multicriteria analysis]. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle kompetenz von lehrkräften. Ergebnisse des forschungsprogramms COACTIV* (pp. 85–113). Waxmann. [CrossRef]

Machts, N., Chernikova, O., Jansen, T., Weidenbusch, M., Fischer, F., & Möller, J. (2024). Categorization of simulated diagnostic situations and the salience of diagnostic information. *Zeitschrift für Pädagogische Psychologie*, *38*(1–2), 3–13. [CrossRef]

Oser, F. (2001). Standards: Kompetenzen von lehrpersonen [Standards: Competencies of teachers]. In F. Oser, & J. Oelkers (Eds.), *Die wirksamkeit der lehrerbildungssysteme. Von der allrounderbildung zur ausbildung professioneller standards* (pp. 215–342). Rüegger.

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275–300. [CrossRef]

Quast, J., Rubach, C., & Lazarides, R. (2021). Lehrkräfteeinschätzungen zu unterrichtsqualität mit digitalen medien: Zusammenhänge zur wahrgenommenen technischen schulausstattung, medienunterstützung, digitalen kompetenzselbsteinschätzungen und wertüberzeugungen [Teachers'assessments of teaching quality with digital media: Correlations with perceived technical school equipment, media support, self-assessment of digital competence, and value beliefs]. *Zeitschrift für Bildungsforschung*, *11*(2), 309–341. [CrossRef]

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(3), 337–386. [CrossRef]

Sailer, M., Maier, R., Berger, S., Kastorff, T., & Stegmann, K. (2024). Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education. *Learning and Individual Differences*, *112*, 102446. [CrossRef]

Schrader, F.-W. (2013). Diagnostische kompetenz von lehrpersonen [Teachers' diagnostic competences]. *Beiträge zur Lehrerbildung*, *31*(2), 154–165. [CrossRef]

Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology*, *60*, 100832. [CrossRef]

Steffens, K. (2006). Self-regulated learning in technology-enhanced learning environments: Lessons of a European peer review. *European Journal of Education*, *41*, 353–379. Available online: https://www.jstor.org/stable/4543062 (accessed on 10 February 2025). [CrossRef]

Stegmann, K. (2020). Effekte digitalen lernens auf den wissens-und kompetenzerwerb in der schule. Eine integration metaanalytischer befunde [Effects of digital media on learning in school. Synthesis of meta-analytical findings]. *Zeitschrift für Pädagogik*, *66*(2), 74–190. [CrossRef]

Süß, H.-M., & Kretzschmar, A. (2018). Impact of cognitive abilities and prior knowledge on complex problem solving performance—Empirical results and a plea for ecologically valid microworlds. *Frontiers in Psychology*, *9*, 629. [CrossRef] [PubMed]

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*, 49. [CrossRef] [PubMed]

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2020). The moderating effect of prior knowledge on the relationship between intelligence and complex problem solving—Testing the Elshout-Raaheim hypothesis. *Intelligence*, *83*, 101502. [CrossRef]

Wekerle, C., Daumiller, M., Janke, S., Dickhäuser, O., Dresel, M., & Kollar, I. (2024). Putting ICAP to the test: How technology-enhanced learning activities are related to cognitive and affective-motivational learning outcomes in higher education. *Scientific Reports*, *14*(1), 16295. [CrossRef] [PubMed]

Wekerle, C., Daumiller, M., & Kollar, I. (2022). Using digital technology to promote higher education learning: The importance of different learning activities and their relations to learning outcomes. *Journal of Research on Technology in Education*, *54*(1), 1–17. [CrossRef]

Willermark, S. (2018). Technological pedagogical and content knowledge: A review of empirical studies published from 2011 to 2016. *Journal of Educational Computing Research*, *56*(3), 315–343. [CrossRef]