The Testing Effect in a Digital Simulation to Foster Pre-Service Teachers' Diagnostic Competence

Lea Grotegut [a], Sebastian T. König [b] and Katrin B. Klingsieck [a]

[a] *Paderborn University, Germany* [b] *Katana Simulations Pty Ltd, Australia*

Submitted to *Educational Psychology*

Author Note

Lea Grotegut – https://orcid.org/0000-0002-8439-8604

Sebastian T. König – https://www.linkedin.com/in/sebkoenig/

A preprint version of this article has been submitted and is to be published as part of the first author's doctoral thesis at Paderborn University, Germany.

The authors report that there are no competing interests to declare.

Correspondence concerning this article should be addressed to Lea Grotegut, Paderborn University, Faculty of Arts and Humanities, Department of Psychology, Educational-psychological assessment and intervention, Warburger Str. 100, 33098 Paderborn, Germany, Email: lea.grotegut@upb.de

Statement of authors' work distribution

LG designed the study and collected and analyzed the data under the supervision of KBK. LG drafted and prepared the manuscript for submission with support from KBK and STK. LG developed the contents of the digital simulation including the quiz. STK programmed the digital simulation and quiz.

Abstract

Digital simulations can train professional competence in a safe and structured environment. Effects of knowledge acquisition as an important aspect of professional competence in digital simulations are, however, inconsistent. The aim of the present study was therefore to explore how knowledge acquisition can be promoted in a digital simulation that trains pre-service-teachers' diagnostic competence. We compared the effects of a quiz in a digital simulation to a mind mapping control condition to investigate the presence and size of a testing effect on knowledge acquisition. Sixty-four pre-service teachers participated in the experimental study, 28 of whom also answered a six-week follow-up questionnaire. Results show greater knowledge acquisition between pretest and posttest in the quiz condition, but not extending to follow-up, indicating a short-term testing effect. Implementing a quiz could therefore be a promising means to enhance knowledge acquisition in digital simulations. However, promoting knowledge retention in digital simulations remains an important task for future research.

Keywords: testing effect, teacher education, knowledge acquisition, diagnostic competence, digital simulation

Introduction

This contribution shows how a quiz embedded in a digital simulation can be used to enhance the acquisition of declarative-conceptual diagnostic knowledge as a prerequisite for pre-service teachers' diagnostic competence (Chernikova, Heitzmann, Fink, et al., 2020; Kramer et al., 2021). Practicing skills in the safe and controllable environment of a digital simulation can be a powerful means for pre-service teachers to develop essential competencies (Badiee & Kaufman, 2014). The digital simulation that was used in the present study has been shown to successfully promote intrinsic motivation and interest (Grotegut & Klingsieck, 2022). However, the simulation's utility for fostering declarative-conceptual diagnostic knowledge has not been demonstrated yet. This is a common problem when it comes to knowledge acquisition via digital simulations (Cant & Cooper, 2010; Kameg et al., 2013). A potential solution for this issue could be using the testing effect, which implies that the active retrieval of previously learned content can improve memorization of this content (Rowland, 2014). Testing seems to be more effective compared to various other learning methods (Roediger & Karpicke, 2006). We implemented a quiz on declarative-conceptual diagnostic knowledge in the digital simulation and tested in an experimental setting whether this quiz does indeed foster knowledge acquisition. Against the backdrop of our results, we will discuss how using a quiz can enhance declarative-conceptual knowledge acquisition in digital simulations. We will first introduce the concept of diagnostic competence and the important role of declarative-conceptual knowledge in competence development before we further elaborate on digital simulations in teacher education and the potential of the testing effect.

Theoretical Framework

Diagnostic Competence

A key aspect of teacher education is the development of professional competencies (Weinert, 2001) such as teaching, counseling, or diagnostic competencies. Diagnostic competence is crucial as teachers' diagnostic judgments build the informational basis for pedagogical decisions. Poor diagnostic and pedagogical decisions can negatively influence students' personal and academic development (Südkamp et al., 2012). For example, misjudging a student's ability could result in denying the student a much-needed intervention. Overall, teachers can judge students' performance with moderate to high accuracy (Südkamp et al., 2012; Urhahne & Wijnia, 2021), but judge students' motivational and emotional characteristics like interest or performance anxiety rather poorly (Spinath, 2005; Urhahne et al., 2010). These empirical findings demonstrate the need to promote diagnostic competence in teacher education.

Correctly judging student characteristics can be seen as the result of a successful diagnostic process based on a well-developed diagnostic competence. We therefore understand diagnostic competence as the "goal-oriented collection and integration of case-specific information to reduce uncertainty in order to make medical or educational decisions" (Heitzmann et al., 2019; p. 4). In addition to person-related variables, such as motivation and personality traits, this process is greatly influenced by diagnostic knowledge (Heitzmann et al., 2019; Kramer et al., 2021). Diagnostic knowledge can be divided into declarative-conceptual knowledge, i.e. knowing concepts and methods related to a certain field (knowing different steps of the diagnostic process), strategic knowledge, i.e. knowing *how* to apply declarative-conceptual knowledge, and conditional knowledge, i.e. knowing under which conditions and why to apply knowledge and strategies (Heitzmann et al., 2018; Stark et al., 2011). Declarative-conceptual knowledge is a prerequisite for strategic knowledge and conditional knowledge. It thus plays a significant role as the foundation for building diagnostic knowledge, which can be applied during diagnostic activities. This hierarchical model is an evolution of earlier models dividing professional knowledge relevant to teaching into three equal kinds of knowledge. In those conceptualizations, professional knowledge is differentiated into content knowledge (CK; knowledge of a subject), pedagogical content knowledge (PCK; knowledge of teaching and learning a subject), and general pedagogical knowledge (GPK; knowledge of teaching and learning in general) (Shulman, 1986, 1987).

The relevance of these three kinds of professional knowledge for diagnostic competence has been demonstrated in several studies: On the outcome level, especially PCK (Kramer et al., 2021) has been found to influence diagnostic judgement accuracy (Ostermann et al., 2018). On the level of the diagnostic process, PCK and GPK are both linked to the application of diagnostic activities, such as interpreting classroom situations (König et al., 2014; Kramer et al., 2021). These findings highlight the potential of diagnostic knowledge to foster diagnostic competence. Combining both conceptualizations of knowledge, declarative-conceptual knowledge entails aspects of CK (e.g., knowing how to calculate the surface area of a rectangle), PCK (e.g., knowledge of common misconceptions regarding the calculation of the surface area of a rectangle), as well as GPK (e.g., knowing how a mathematics lesson should be structured to effectively address student misconceptions). As declarative-conceptual knowledge is considered a prerequisite for strategic and conditional knowledge, focusing on building a strong foundation of declarative-conceptual diagnostic knowledge (DCDK) early on in teacher education seems a promising approach in the development of diagnostic competence. In this context, the present study focuses on promoting DCDK as a part of teachers' diagnostic competence.

Diagnostic Knowledge and the Need for Knowledge Application

It is crucial to offer pre-service teachers the opportunity to *apply* DCDK and to practice pedagogical actions to build strategic and conditional diagnostic knowledge and professional competence (Hascher et al., 2004). Practicing pedagogical actions in school, however, can create problems for pre-service teachers as well as students. The sudden complexity of the school environment, the change of role from student to teacher, and the gap between theory and practice can impose a feeling of being overwhelmed on pre-service teachers (Heitzmann et al., 2019). Moreover, invalid diagnostic and pedagogical decisions can pose risks for students' personal and academic development (Südkamp et al., 2012). Due to these high stakes, repeated practice in real-life situations is usually difficult to implement in pre-service teacher education. Hence, *approximations to practice* can be especially useful for practicing pedagogical actions in authentic learning situations whilst decomposing these situations into segments suited to the learner's proficiency (Grossman et al., 2009). Digital simulations are a promising example of such approximations to practice and can offer a tool to apply pedagogical actions in a safe and controllable environment.

Digital Simulations in Teacher Education

A digital simulation is a simplified representation of a real system, such as a school, that includes the components of the real system at physical (e. g., the visual representation of a classroom) and informational levels (e. g., information about student performance). The components can be influenced by the person interacting with the digital simulation (Sauvé et al., 2007). A major advantage of digital simulations in learning settings is the option to go through realistic situations multiple times under controllable circumstances (Badiee & Kaufman, 2014) without having to fear negative consequences (Dieker et al., 2014). A meta-analysis found digital and non-digital simulations to be well-suited to promote complex skills in higher education (Chernikova, Heitzmann, Stadler, et al., 2020). They can also be motivating, especially on an intrinsic level (Grotegut & Klingsieck, 2022; Stavroulia & Lanitis, 2017). This seems particularly important considering that intrinsic motivation is generally associated with better learning outcomes and learners' well-being (Bailey & Phillips, 2016; Vansteenkiste et al., 2004). Despite these pertinent advantages of digital simulations, it seems to be difficult to promote declarative-conceptual knowledge acquisition via digital simulations. In fact, digital simulations' effects on knowledge acquisition and retention are sometimes positive (Gebreheat et al., 2022), sometimes mixed (Cant & Cooper, 2010), or even nonexistent (Grotegut & Klingsieck, 2022; Kameg et al., 2013). Positive effects might be heavily influenced by publication bias in digital simulation research (Sitzmann, 2011) and the mixed findings question the extent to which digital simulations can actually enhance knowledge acquisition compared to studying via reading texts or lectures. In contrast, quizzes embedded in digital simulations leverage the so-called testing effect and have the potential to promote practical skills while also fostering knowledge acquisition (Kantar et al., 2020).

The Testing Effect

The testing effect refers to "the finding that retrieving information from memory can, under many circumstances, strengthen one's memory of the retrieved information" (Rowland, 2014, p. 1). In practice, learners engage in a previously unknown topic (initial learning phase) followed by a testing phase in which learners are tested on the learned content, for example with multiple-choice questions (Burdo & O'Dwyer, 2015). In addition, learners can receive feedback on the correctness of their answers after each question or after the whole test (Butler & Roediger, 2008). Retention of the learning content is then assessed in a final test (Yang et al., 2021). The testing effect has been found to be of medium to large effect size in several meta-analyses in experimental (Rowland, 2014) as well as classroom settings (Yang et al., 2021), and a combination of both (Adesope et al., 2017). Including feedback on each answer during the initial testing phase has been found to enhance the testing effect and result in better learning outcomes by preventing the memorization of false answers, especially in multiple-choice tests (Roediger & Butler, 2011; Vojdanoska et al., 2010). What makes the testing effect so appealing is that it results in better knowledge acquisition compared to various control conditions such as note-taking (Heitmann et al., 2018) and restudying (Kirk-Johnson et al., 2019; Roediger & Karpicke, 2006), but also compared to control conditions aiming at higher taxonomy levels such as concept mapping (Karpicke & Blunt, 2011). Moreover, it has been found to promote knowledge acquisition, problem-solving (Yang et al., 2021), and skill acquisition (Kromann et al., 2010). Although the testing effect has shown to be applicable to various learning contexts, topics, and learning outcomes, research has been paying less attention to the impact of different media and materials used during the initial learning phase. In most studies, participants learn a new topic via reading a text before the testing phase, but the testing effect could also be applied to multimedia learning (Johnson & Mayer, 2009), where learners acquired the necessary knowledge by watching a narrated animation. The testing effect has, however, not yet been applied to digital, interactive learning environments such as educational games or simulations, which become increasingly important as education embraces digitalization (Zhonggen, 2019). There is also an ongoing discussion about the robustness of the testing effect with more complex learning materials (e.g., materials with high interactivity). A literature review argues that studies which find a testing effect often rely on almost literal recall of the learning content, a non-restudy control condition (hampering recall in the control group), and include feedback during the testing phase, thereby enhancing the effects found in favor of the experimental group. On the contrary, studies including complex learning materials and/or demanding transfer questions do not find a testing effect (van Gog & Sweller, 2015). However, several counterexamples have shown that the testing effect does indeed hold for complex learning materials such as mechanical devices (McDaniel et al., 2009) or maps (Rohrer et al., 2010) and can also positively influence recall of non-tested material (Chan, 2009). The present study is, to our knowledge, the first instance of a complex, highly interactive digital simulation being used to investigate the testing effect.

The Present Study

The aim of the present study is to investigate whether the testing effect can be used in a digital simulation to promote the acquisition of DCDK in pre-service teachers. We hypothesized that a quiz implemented in a digital simulation results in greater acquisition of DCDK than the control condition. We tested our hypothesis in an experimental setting where participants were randomly assigned to either the so-called quizzing condition or a control condition. Both groups worked with the same digital simulation. While the learners in the quizzing condition completed a test on the learning content embedded in the digital simulation, learners in the control condition were assigned an alternative task. In addition to DCDK, we explored learners' interest as an indicator of intrinsic motivation, cognitive load as indicator on how cognitively demanding the simulation is, and positive and negative emotions as indicators for how students felt when working with the digital simulation. The results of this study will contribute to instructional science by a) investigating whether a testing effect can be found if the study material is highly interactive and complex as in a digital simulation. As the testing effect thus far has only been applied to learning materials in the form of texts, animations, images, and videos, this study broadens the research in this field by using a highly interactive, digital simulation as an authentic learning environment. Moreover, this study contributes to digital simulation research by b) adapting a digital simulation for pre-service teachers by adding a quiz to enhance declarative-conceptual knowledge acquisition and, thus, advance the scope of professional competence development.

Method

The study was conducted at the University of Paderborn with teacher education students from different courses of study and included three points of measurement: immediately before participants began working with the digital simulation (T1), immediately after participants completed the quiz or the alternative task (T2), and a follow-up approximately six weeks after the intervention. Participants filled out an online survey at each point of measurement. The survey was sent out to their e-mail addresses via personal link. All participants took part voluntarily and gave their informed consent before the study. Data were collected anonymously.

Sample

We recruited participants via various public channels at our university, including Facebook groups, Instagram, and e-mail distribution lists. We also advertised the experiment in different teacher education seminars. Participants had to be enrolled in a teacher education program and were offered €10 and a certificate of participation as incentives. In total, 98 pre-service teachers enrolled in the experiment via a short online questionnaire. Of the 98 registered participants, 66 ultimately took part in the experiment. Twenty-eight of those additionally answered the online follow-up (T3) questionnaire, resulting in a total sample of n = 64 (as not all participants answered every question) for the first (T1) and second point of measurement (T2) and n = 28 for all three points of measurement. Descriptive data for both samples can be found in Table 1. There were no group differences between participants in the T1-T2-sample and participants in the T1-T2-T3-sample regarding demographic variables and prior knowledge (Table 2), meaning that there was no indication of a systematic dropout after the post-test. Participants were randomly assigned to either the quizzing (experimental group, EG, $n_{T1T2} = 33$; $n_{T1T2T3} = 15$) or the mind mapping condition (control group, CG, $n_{T1T2} = 31$; $n_{T1T2T3} = 13$).

There were no differences between the experimental and the control group regarding age, sex, semesters enrolled, prior DCDK, positive emotions at T1, or negative emotions at T1 (Table 3).

Table 1

Samples

Variable	Sample T1 and T2 (n = 64)						Sample T1, T2 and T3 (n = 26)					
	EG			CG			EG			CG		
	M	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
Age	22.7	2.4	19-30	23.48	4.77	18-45	22.57	2.59	19-30	24.38	6.82	18-45
GPA (Abitur grade)	2.35	.53	1.4-3.3	2.27	.66	1.0-3.2	2.35	.43	1.7-3.3	2.38	.68	1.0-3.1
Semesters enrolled	4.64	2.71	1-9	4.16	2.51	1-9	5.07	2.7	1-9	4.00	2.52	1-8

Table 2

Dropout Analysis (ANOVA)

Variable	T1 and T2 only (n = 64)		T1, T2 and T3 (n = 26)		*F*	*df, err. df*	*p*	η_p^2
	M	*SD*	*M*	*SD*				
Age	22.71	2.5	23.62	5.03	.907	1, 62	.345	.003
GPA (Abitur grade)	2.27	.63	2.37	.54	.386	1, 62	.537	.006
Semesters enrolled	4.29	2.62	4.58	2.63	.186	1, 62	.668	.003
Prior declarative-conceptual diagnostic knowledge	53.26	7.31	53.73	6.94	.066	1, 62	.799	.001
Positive emotions	2.77	.7	2.88	.71	.379	1, 62	.541	.006
Negative emotions	1.97	.84	1.95	.74	.009	1, 62	.923	.000
	%		%					
Women	84.2		92.3		.178	1, 62	.67	.014

Table 3

Group Comparisons Between Experimental Group and Control Group at Pretest (ANOVA)

Variable	Sample T1 and T2 (n = 64)				Sample T1, T2 and T3 (n = 26)			
	F	*df, err. df*	*p*	η_p^2	*F*	*df, err. df*	*p*	η_p^2
Age	.708	1, 62	.403	.011	1.324	1, 24	.261	.052
Sex	1.091	1, 62	.3	.017	2.281	1, 24	.144	.087
GPA (Abitur grade)	.292	1, 62	.591	.005	.023	1, 24	.88	.001
Semesters enrolled	.528	1, 62	.47	.008	1.079	1, 24	.309	.043
Prior declarative-conceptual diagnostic knowledge	.972	1, 62	.328	.015	.005	1, 24	.946	< .001
Positive emotions	.016	1, 62	.901	< .001	.378	1, 24	.544	.016
Negative emotions	1.875	1, 62	.176	.029	.753	1, 24	.394	.03

The Digital Simulation

We have developed the digital simulation as a tool to be used in seminars focusing on assessment and evaluation for teachers. The main purpose of the simulation is to provide learners with an opportunity to practice their diagnostic skills in a realistic setting without having to fear actual pedagogical consequences and thereby support them in developing their diagnostic competence.

The digital simulation presents a virtual school environment where users take the role of a teacher in a sixth grade (eleven- to twelve-year-old students). The focus of the digital simulation is collecting and systematizing students' information, such as grades, behavior, family situation, strengths, and difficulties. The aim is for users to make a well-founded diagnostic decision on virtual students' learning predispositions. Users can observe student behavior in a virtual lesson, gather information on students from grade reports, teacher notes, or workbooks, and have conversations with students, parents, or colleagues. Some virtual students present different distinctive behavioral, emotional, or cognitive characteristics. For the present study, users worked on a specific virtual student presenting aspects of attention difficulties and symptoms of attention deficit hyperactivity disorder (ADHD). The tasks integrated in the digital simulation consist of ten diagnostic steps adapted to the diagnostic process (Fischer et al., 2014; Wildgans-Lang et al., 2020): 1) behavioral observation, 2) collecting information on student performance, 3) systematization of information, 4) generating working hypotheses, 5) gaining theoretical knowledge on relevant topics, 6) adapting hypotheses, 7) planning necessary next steps, 8) evaluating and interpreting collected data (e.g. test results provided in the digital simulation), 9) evaluating hypotheses, and 10) developing a tentative diagnosis and practical implications. Users work on every diagnostic step in detail and are provided with relevant literature to assess the virtual students' learning predispositions. The ten diagnostic steps are designed to be completed in approximately two hours. The simulation is accessible via most common web browsers.

Experimental Condition: The Quiz in the Digital Simulation

We aimed at enhancing the testing effect with feedback (cf. Roediger & Butler, 2011) by implementing a quiz including feedback after each question. The quiz was designed to fit into the narrative of the digital simulation and thus was embedded in an e-mail between the user's character, a sixth-grade teacher, and an acquainted child and adolescent psychotherapist. The quiz was only unlocked after users completed an essential diagnostic step in the digital simulation so as not to allow them to take the quiz before working on the diagnostic steps and acquiring the DCDK necessary to answer the quiz questions. It consisted of nine multiple choice and six short answer questions. The items focused on DCDK about assessment and evaluation for teachers in general as well as symptoms and background knowledge of ADHD, the diagnostic process, and support measures in the case of ADHD. After answering a question, users were provided with feedback telling them whether their answer was correct or incorrect. The rationale for the correct answer was then included in the e-mail following the question as part of the conversation between the teacher and the psychotherapist. Users had the opportunity to go back to previous e-mails, questions, and answers at any point during the quiz.

For this study, we used two versions of the digital simulation: one that included the quiz (EG) and one that did not (CG). Only users assigned to the EG could access the quiz. Users in the CG were asked to create a mind map ("Now design a mind map in which you systematically present the knowledge and information you have acquired on the topic of *ADHD – theoretical background, assessment, and support*."). Designing the mind map took approximately the same amount of time as taking the quiz.

Measures

Declarative-Conceptual Diagnostic Knowledge (DCDK)

We assessed DCDK at all three points of measurement with 24 multiple choice items. The items had different numbers of answer options out of which either one or multiple answers were correct. Of these, 15 items were taken from a large existing item pool which has been established in teacher education at our university and tested over several years in our research. The items covered different topics regarding DCDK, including the concept of professional diagnostics and its relation to intuitive diagnostics ('What distinguishes systematic diagnostics from intuitive, non-professional diagnostics?'), aspects relevant to the diagnosis of ADHD, e.g. psychometric tests ('Which of these psychometric tests is used to assess intelligence in children and adolescents?'), and the concept of neurodevelopmental disorders and the relationship between different symptoms of ADHD and attention deficit in general ('Which of the following are symptoms of ADHD?'). The other nine items were taken from the quiz in the digital simulation. They covered different diagnostic concepts relevant to the diagnosis of ADHD. These included diagnostic criteria of ADHD, the diagnosis of ADHD and the teacher's role in the diagnostic process ('Who may diagnose ADHD?'), and the relationship between diagnosis and student support as well as knowledge about specific means of supporting students with ADHD and attention deficits ('How does a token system work?'). In addition, the diagnostic process, its relation to pedagogical support measures, and the teacher's role in this process were covered ('As a teacher, what steps can you take to support a child with ADHD?'). As strategic and conditional knowledge can only be built after acquiring declarative-conceptual knowledge and learners in our sample did not have any relevant prior knowledge, we focused on DCDK acquisition and refrained from assessing strategic and conditional knowledge in the present study. Item difficulty at T1 was between .28 and .89.

Interest

We assessed interest directly after learners completed their interaction with the digital simulation using a subscale of the *questionnaire to assess current motivation in learning situations* (FAM, Rheinberg et al., 2001). The scale includes four items scored on a 7-point Likert scale ranging from *does not apply* to *applies* (e.g., 'After reading the instruction, the task seemed very interesting to me'). Internal consistency was $\alpha = .89$.

Cognitive Load

We used four items of an instrument developed by Klepsch et al. (2017) including the subscales *intrinsic cognitive load* (two items, e.g., 'Working with the digital simulation was very complex'; $\alpha = .63$) and *extraneous cognitive load*, (two items, e.g., 'When working with digital simulation, it is tedious to identify the key information'; $\alpha = .65$), to assess learners' cognitive load.

Emotions

Participants' positive and negative emotions were assessed across all three points of measurement. Participants were presented six positive and five negative emotions from the German version of the *Positive and Negative Affect Schedule* (PANAS) (Breyer & Bluemke, 2016; Watson et al., 1988) and asked 'How often did you experience the following emotions during the past semester?'. Each emotion (e.g., 'pride', 'frustration') was to be rated on a five-point-scale from *never* to *very often*. Internal consistency was $\alpha_{T1} = .77$, $\alpha_{T2} = .76$, and $\alpha_{T3} = .85$ for positive emotions, and $\alpha_{T1} = .76$, $\alpha_{T2} = .74$, and $\alpha_{T3} = .68$ for negative emotions.

Data Analyses

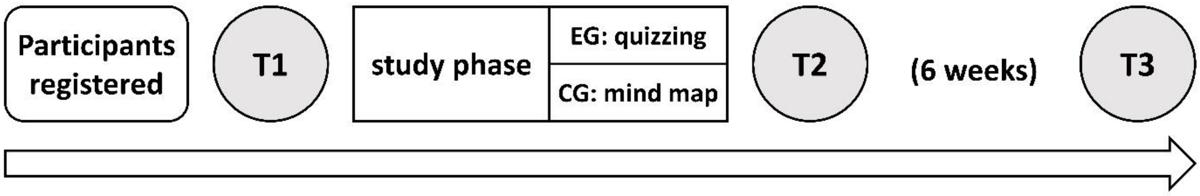
We used an analysis of covariance with repeated measurements (ANOVA) to determine the development of DCDK dependent on the two experimental conditions. A multivariate analysis of covariance (MANOVA) with repeated measurements was performed for positive and negative emotions. We analyzed interest and cognitive load, which were only assessed at T2, by two separate MANOVA. Effect sizes were calculated using partial eta squared η_p^2. Statistical analyses were performed in SPSS version 28.0.

Procedure

Participants registered to take part in the experiment at one of four different dates within two weeks. The experiment took three hours. All sessions were held by the same instructor and were identically structured: Participants took a seat at a prepared desk where they found the working materials as well as a sign indicating to which of the two groups they were randomly assigned. Participants were welcomed before they filled out the first online survey. Afterwards, they received a short introduction to the digital simulation before they started working individually on the assignments in the digital simulation with their own laptops or a university laptop. All participants had the same amount of time to complete each assignment in the digital simulation as the instructor gave the signal to move on to the next assignment after a certain amount of time. Participants were provided with standardized work sheets to record their results. They also received a handout with information on the assessments and interventions for ADHD and symptoms of attention deficit. This handout was to mimic the literature that students usually access when working with the simulation. It was relevant to certain assignments in the digital simulation and participants were instructed to use the handout, which was placed face-down on their desk, when working on these assignments only. After they completed the assignments in the digital simulation, the instructor collected the worksheets and handouts. Participants were then instructed to flip the last sheet that was placed face-down on their desk and read the assignment, which included instructions for either the quiz in the digital simulation (EG) or instructions and a worksheet for the mind mapping task (CG). While participants were completing the quiz or mind map, the instructor ensured that no additional resources were used to complete the tasks. After they completed the quiz or mind map, all participants filled out the second online survey. Approximately six weeks after the session, participants received a link to fill out the follow-up online survey (for an overview of the study procedure see Figure 1).

Figure 1

Course of the Experiment



Results

Means and standard deviations are summarized in Table 4. All ANOVA and MANOVA results are summarized in Table 5 and described in more detail in the following section, structured according to dependent variables.

Diagnostic Knowledge

Addressing our hypothesis, we investigated the effects of the quizzing and mind mapping condition on DCDK acquisition with a repeated measures ANOVA. There was a significant large main effect for time (n_{T1T2}: $\eta_p^2 = .66$; $n_{T1T2T3:}$ $\eta_p^2 = .62$), indicating that both the EG and the CG gained DCDK while engaging in the digital simulation. Moreover, there was a significant medium to large interaction effect (n_{T1T2}: $\eta_p^2 = .09$; $n_{T1T2T3:}$ $\eta_p^2 = .19$) for the EG, confirming our hypothesis. Thus, the EG acquired more DCDK during the intervention than the CG, demonstrating that the quiz was an effective tool to boost DCDK acquisition when learning in the digital simulation. This advantage for the EG was evident for increase in DCDK between T1 and T2, but not found in the smaller sample of all three points of measurement: While the level of DCDK in the CG remained the same between T2 and T3, it slightly decreased in the EG, as an interaction effect shows ($p = .021$, $n_{T1T2T3:}$ $\eta_p^2 = .20$).

Interest

MANOVA results showed no significant multivariate or univariate effects for interest, which was measured directly after completing the quizzing or mind mapping task.

Cognitive Load

MANOVA results showed no significant multivariate effect for cognitive load, which included the subscale *Intrinsic Cognitive Load* and two items of the *Extraneous Cognitive Load* subscale (see methods section). There were also no univariate effects for either of the two subscales.

Emotions

We found a significant multivariate effect for time, showing that negative emotions decreased from T1 to T2 but were significantly higher at T3 than at either T1 or T2 ($n_{T1T2:}$ η_p^2 = .34 $n_{T1T2T3:}$ η_p^2 = .64), and positive emotions were significantly higher at T3 than at T1 ($n_{T1T2T3:}$ η_p^2 = .21), as indicated by large effect sizes. There was no significant multivariate interaction between time and group for positive and negative emotions and no significant multivariate group effect. Thus, both tasks have led to a decrease in negative affect in the short term.

Table 4

Means and Standard Deviations

Variable	T1(n = 64)		T2 (n = 64)		T3 (n = 26)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Declarative-conceptual diagnostic knowledge						
EG	54.30	7.92	65.24	8.02	62.71	5.66
CG	52.55	6.15	59.29	6.21	60.08	5.42
Positive emotions						
EG	2.82	0.73	2.82	0.82	3.29	0.86
CG	2.80	0.69	3.13	0.59	3.14	0.57
Negative emotions						
EG	2.10	0.98	1.58	0.66	2.63	0.67
CG	1.83	0.52	1.39	0.51	2.70	0.70
Interest [a]						
EG	–	–	5.47	1.26	–	–
CG	–	–	5.70	0.96	–	–
Intrinsic cog. load [a]						
EG	–	–	4.12	1.21	–	–
CG	–	–	4.18	1.45	–	–
Extraneous cog. load [a]						
EG	–	–	2.88	1.33	–	–
CG	–	–	3.05	1.16	–	–

Note. EG = experimental group, CG = control group.

[a] Interest and cognitive load were assessed at T2 only.

Table 5

ANOVA and MANOVA

Variable in both samples	time				group				time x group			
	F	*df, err. df*	*p*	η_p^2	*F*	*df, err. df*	*p*	η_p^2	*F*	*df, err. df*	*p*	η_p^2
Sample T1-T2 (n = 64)												
Declarative-conceptual diagnostic knowledge	119.00	1, 62	< .001	.66	5.69	1, 62	.02	.08	6.33	1, 62	.01	.09
Positive emotions	2.50	1, 62	0.12	.04	2.32	1, 62	.33	.02	2.66	1, 62	.11	.04
Negative emotions	31.44	1, 62	< .001	.34	2.32	1, 62	.13	.04	0.20	1, 62	.66	.00
Interest	–	–	–	–	.71	1, 62	.40	.01	–	–	–	–
Intrinsic cog. load	–	–	–	–	.03	1, 62	.87	.00	–	–	–	–
Extraneous cog. load	–	–	–	–	.34	1, 62	.56	.05	–	–	–	–
Sample T1-T2-T3 (n = 26)												
Declarative-conceptual diagnostic knowledge	39.63	2, 48	< .001	.62	2.13	1, 24	.16	.08	5.48	2, 48	.01	.19
Positive emotions	2.03	1.76, 42.27	.15	.08	.05	1, 24	.83	.00	.51	1.76, 42.27	.58	.02
Negative emotions	42.49	2, 48	< .001	.64	.54	1, 24	.47	.02	.97	2, 48	.39	.04

Discussion

In the present study, we investigated whether the testing effect can be utilized in a digital simulation to promote DCDK acquisition in pre-service teachers. We found increased DCDK acquisition for the EG, which completed a quiz, compared to the CG, which created a mind map.

In detail, both groups showed a significant increase in DCDK after learning in the digital simulation. The EG gained significantly more knowledge between the pre- and posttest compared to the CG, confirming our hypothesis, which assumed a larger increase in DCDK for the EG compared to the CG. Therefore, the quiz has shown to be an effective tool to enhance the digital simulation regarding its effects on declarative-conceptual knowledge acquisition. Further exploration with a smaller subsample showed that this effect, however, did not seem to last at 6-week-follow-up. Negative emotions decreased in both groups during the experiment, indicating that using the digital simulation lowered learners' negative affect. Positive emotions were higher at follow-up compared to pretest in both groups, indicating an increase in positive affect over the course of the semester. We found no group differences regarding interest, intrinsic or extraneous cognitive load.

Limitations

Some limitations regarding the study design and interpretation of the results need to be addressed. The sample across all three points of measurement was relatively small with n = 28 learners as more than half of the participants dropped out after T2. This indicates that the decrease in knowledge from T2 to T3 needs to be interpreted with caution. However, results obtained from the T1-T2-T3 sample can be seen as an additional confirmation of the results obtained from the T1-T2 sample: In the T1-T2-T3 sample, we also found a significant large effect in favor of the EG in terms of DCDK acquisition between pre- and posttest.

Furthermore, our sample consisted of pre-service teachers from the same university and, therefore, cannot be considered representative of pre-service teachers in general. However, pre-service teachers who took part in this experiment covered a fairly large range of age, semesters enrolled, and study programs, indicating that this small sample was nevertheless able to depict a diverse group of students.

Concerning the experimental conditions, the initial learning phase during which learners engaged in the digital simulation could only partly be standardized. Learners had the same amount of time and were provided with the same materials for the tasks at hand. However, it could not be assured that they made use of exactly the same content in the digital simulation to solve the tasks. The simulation includes a wide range of learning materials for learners to engage with individually and thereby differs from standardized learning materials such as texts or videos. Very standardized materials, however, tend to not be able to depict authentic learning contexts as learners in higher education often need to synthesize different sources of information and media in their studies. Moreover, as the digital simulation offers many advantages for teacher education (Kaufman & Ireland, 2016), it should not be neglected as a study condition for investigating the testing effect.

Concerning the realization of the testing effect, the study design included elaborate feedback for the quiz group's answers, which is expected to enhance learning and retention (Rowland, 2014). When creating the quiz, it was a deliberate design decision to promote DCDK acquisition by providing feedback. Though it could be argued that the EG's access to more knowledge (in the form of feedback) gave them a distinct advantage over the CG, which did not receive feedback (van Gog & Sweller, 2015). Also, the CG did not have access to the learning materials while creating the mind map as it would be the case in a restudy condition. The entire digital simulation was too complex a learning tool to be made available during the control condition's mind map task. These issues are also pointed out by van Gog and Sweller (2015), who state that a true testing effect does not occur when complex learning materials are used during the initial learning phase. Thus, providing feedback in the experimental condition could have confounded the mere effect of the quiz. However, by including feedback we have designed a version of the quiz that has shown to be effective and is in line with studies that demonstrated an enhanced testing effect when using feedback (e.g., Butler et al., 2008; Butler & Roediger, 2008; Vojdanoska et al., 2010). Moreover, other studies were able to find a testing effect even when testing is compared to open-book concept mapping where participants in the control group could access the initial learning materials while creating the concept map (Burdo & O'Dwyer, 2015; Karpicke & Blunt, 2011). As these control groups did not outperform the experimental groups, it can be assumed that the effect we found in favor of the EG is not – at least fully – due to the provision of additional knowledge resources in the form of feedback, but to testing itself. Although a comparison of the digital simulation quizzing condition to a control condition with similar access to the learning content would be worth investigating in future research, we consider the mind mapping task authentic for educational settings and thus our results relevant, especially for applied research regarding the testing effect, which has received rather little attention so far.

Contributions and Outlook

Overall, the results of this study contribute to instructional science in two ways. First, we demonstrated that a digital simulation can promote DCDK acquisition in pre-service teachers. Digital simulation research is still in its early stages regarding the investigation of effects on knowledge acquisition, especially in experimental study designs. The present study helps to enrich this research by investigating the effects of a digital simulation in an experimental, randomized pre-post-follow-up control group design. Moreover, the only difference between the simulation used for EG and CG was the availability of the quiz. This allows us to isolate the effects of the quiz on learning rather than comparing the digital simulation to entirely different learning settings (Imlig-Iten & Petko, 2018).

Second, the results show that a testing effect can be found also if the study material is presented in a highly interactive digital simulation rather than in the form of more basic learning materials like texts, images, or videos. Up until this point, only few studies investigated and found a testing effect with multimedia learning materials (Johnson & Mayer, 2009). Also, gamified tests can have advantages for higher performing learners but result in impaired retention (Sanchez et al., 2020). Our results add to this research by showing that a quiz implemented in an interactive, digital simulation can enhance DCDK acquisition compared to a control group. We found evidence that this effect might, however, not last for longer time periods. In this regard, our results are similar to those from Sanchez et al. (2020). Incorporating more complex question types into the quiz, such as free recall questions (Blunt & Karpicke, 2014), or adapting the difficulty level of the questions to the learners' ability level (Heitmann et al., 2018) could be ways to improve knowledge retention while learning in a digital simulation and thereby further contributing to digital simulation as well as testing effect research.

Acknowledgements

We thank Lea Klinge for her contribution in conceptualizing and designing the quiz in the digital simulation.

Declaration of Interest Statement

The authors report that there are no competing interests to declare.

Disclosure Statement

A preprint version of this article has been submitted and is to be published as part of the first author's doctoral thesis at Paderborn University, Germany.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Badiee, F., & Kaufman, D. (2014). Effectiveness of an online simulation for teacher education. *Journal of Technology and Teacher Education*, *22*(2), 167–186. https://www.learntechlib.org/p/45934/

Bailey, T. H., & Phillips, L. J. (2016). The influence of motivation and adaptation on students' subjective well-being, meaning in life and academic performance. *Higher Education Research & Development*, *35*(2), 201–216. https://doi.org/10.1080/07294360.2015.1087474

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*(3), 849–858. https://doi.org/10.1037/a0035934

Breyer, B., & Bluemke, M. (2016*). Deutsche Version der positive and negative affect schedule PANAS (GESIS Panel)* [German version of the positive and negative affect schedule PANAS]. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/ZIS242

Burdo, J., & O'Dwyer, L. (2015). The effectiveness of concept mapping and retrieval practice as learning strategies in an undergraduate physiology course. *Advances in Physiology Education*, *39*(4), 335–340. https://doi.org/10.1152/advan.00041.2015

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. https://doi.org/10.3758/MC.36.3.604

Cant, R. P., & Cooper, S. J. (2010). Simulation-based learning in nurse education: Systematic review. *Journal of Advanced Nursing*, *66*(1), 3–15. https://doi.org/10.1111/j.1365-2648.2009.05240.x

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*(2), 153–170. https://doi.org/10.1016/j.jml.2009.04.004

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

Dieker, L. A., Rodriguez, J. A., Lignugaris/Kraft, B., Hynes, M. C., & Hughes, C. E. (2014). The potential of simulated environments in teacher education. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *37*(1), 21–33. https://doi.org/10.1177/0888406413512683

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://doi.org/10.14786/flr.v2i2.96

Gebreheat, G., Whitehorn, L. J., & Paterson, R. E. (2022). Effectiveness of digital simulation on student nurses' knowledge and confidence: An integrative literature review. *Advances in Medical Education and Practice*, *13*, 765–775. https://doi.org/10.2147/AMEP.S366495

Grossman, P., 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

Grotegut, L., & Klingsieck, K. B. (2022). Wie können unterschiedliche Aspekte diagnostischer Kompetenz gefördert werden? Drei Maßnahmen im Vergleich [How can different aspects of diagnostic competence be promoted? A comparison of three measures]. *Zeitschrift Für Pädagogische Psychologie*, Article 1010-0652/a000352. Advance online publication. https://doi.org/10.1024/1010-0652/a000352

Hascher, T., Cocard, Y., & Moser, P. (2004). Forget about theory - practice is all? Student teachers' learning in practicum. *Teachers and Teaching*, *10*(6), 623–637. https://doi.org/10.1080/1354060042000304800

Heitmann, S., Grund, A., Berthold, K., Fries, S., & Roelle, J. (2018). Testing is more desirable when it is adaptive and still desirable when compared to note-taking. *Frontiers in Psychology*, *9*, Article 2596. https://doi.org/10.3389/fpsyg.2018.02596

Heitzmann, N., Fischer, F., & Fischer, M. R. (2018). Worked examples with errors: When self-explanation prompts hinder learning of teachers diagnostic competences on problem-based learning. *Instructional Science*, *46*(2), 245–271. https://doi.org/10.1007/s11251-017-9432-2

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M., Ufer, S., Schmidmaier, R., Neuhaus, B., & Siebeck, M. (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.

Imlig-Iten, N., & Petko, D. (2018). Comparing serious games and educational simulations: Effects on enjoyment, deep thinking, interest and cognitive learning gains. *Simulation & Gaming*, *49*(4), 401–422. https://doi.org/10.1177/1046878118779088

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, *101*(3), 621–629. https://doi.org/10.1037/a0015183

Kameg, K. M., Englert, N. C., Howard, V. M., & Perozzi, K. J. (2013). Fusion of psychiatric and medical high fidelity patient simulation scenarios: Effect on nursing student knowledge, retention of knowledge, and perception. *Issues in Mental Health Nursing*, *34*(12), 892–900. https://doi.org/10.3109/01612840.2013.854543

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science (New York, N.Y.)*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Kaufman, D., & Ireland, A. (2016). Enhancing teacher education with simulations. *TechTrends*, *60*(3), 260–267. https://doi.org/10.1007/s11528-016-0049-0

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, *115*, Article 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

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

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

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

Kromann, C. B., Bohnstedt, C., Jensen, M. L., & Ringsted, C. (2010). The testing effect on skills learning might last 6 months. *Advances in Health Sciences Education*, *15*(3), 395–401. https://doi.org/10.1007/s10459-009-9207-x

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, *20*(4), 516–522. https://doi.org/10.1111/j.1467-9280.2009.02325.x

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

Roediger, H. L., III., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(1), 233–239. https://doi.org/10.1037/a0017678

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Sanchez, D. R., Langer, M., & Kaur, R. (2020). Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Computers & Education*, *144*, Article 103666. https://doi.org/10.1016/j.compedu.2019.103666

Sauvé, L., Renaud, L., Kaufman, D., & Marquis, J.-S. (2007). Distinguishing between games and simulations: A systematic review. *Educational Technology & Society*, *10*(3), 247–256.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–23.

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

Stark, R., Kopp, V., & Fischer, M. R. (2011). Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. *Learning and Instruction*, *21*(1), 22–33. https://doi.org/10.1016/j.learninstruc.2009.10.001

Stavroulia, K.-E., & Lanitis, A. (2017). On the potential of using virtual reality for teacher education. In P. Zaphiris & A. Ioannou (Eds.), *Lecture notes in computer science: Vol. 10295. Novel learning ecosystems* (Vol. 10295, pp. 173–186). Springer. https://doi.org/10.1007/978-3-319-58509-3_15

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.

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

Urhahne, D., Zhou, J., Stobbe, M., Chao, S.-H., Zhu, M., & Shi, J. (2010). Motivationale und affektive Merkmale unterschätzter Schüler [Motivational and affectice characteristics of underestimated students]. *Zeitschrift Für Pädagogische Psychologie*, *24*(3-4), 275–288. https://doi.org/10.1024/1010-0652/a000021

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*(2), 247–264. https://doi.org/10.1007/s10648-015-9310-x

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology*, *87*(2), 246–260.

Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*(8), 1183–1195. https://doi.org/10.1002/acp.1630

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. Hersh Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–66). Hogrefe & Huber Publishers.

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

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

Zhonggen, Y. (2019). A meta-analysis of use of serious games in education over a decade. *International Journal of Computer Games Technology*, *2019*, 1–8. https://doi.org/10.1155/2019/479703