Promoting diagnostic reasoning in teacher education: the role of case format and perceived authenticity

Sarah Bichler[1,2] · Michael Sailer[3] · Elisabeth Bauer[3] · Jan Kiesewetter[4] ·
Hanna Härtl[2] · Martin R. Fischer[4] · Frank Fischer[2]

Abstract

Teachers routinely observe and interpret student behavior to make judgements about whether and how to support their students' learning. Simulated cases can help pre-service teachers to gain this skill of diagnostic reasoning. With 118 pre-service teachers, we tested whether participants rate simulated cases presented in a serial-cue case format as more authentic and become more involved with the materials compared to cases presented in a whole case format. We further investigated whether participants with varying prior conceptual knowledge (what are symptoms of ADHD and dyslexia) gain more strategic knowledge (how to detect ADHD and dyslexia) with a serial-cue versus whole case format. We found that the case format did not impact authenticity ratings but that learners reported higher involvement in the serial-cue case format condition. Bayes factors provide moderate evidence for the absence of a case format effect on strategic knowledge and strong evidence for the absence of an interaction of case format and prior knowledge. We recommend using serial-cue case formats in simulations as they are a more authentic representation of the diagnostic reasoning process and cognitively involve learners. We call for replications to gather more evidence for the impact of case format on knowledge acquisition. We suggest a further inquiry into the relationship of case format, involvement, and authenticity but think that a productive way forward for designing authentic simulations is attention to aspects that make serial-cue cases effective for diverse learners. For example, adaptive feedback or targeted practice of specific parts of diagnostic reasoning such as weighing evidence.

Keywords Authenticity · Involvement · Case format · Simulation-based learning · Diagnostic reasoning

Teachers constantly observe students' behavior, evaluate students' work, and track students' learning progress. This diagnosis of common patterns of student thinking and development is a core practice of teaching (e.g., Grossman, 2021). To master this practice and best support all their students, teachers need to know how to notice and interpret signs of understanding, learning, frustration, success, or struggle (e.g., Robertson et al., 2016;

Sarah Bichler and Michael Sailer share first authorship of this manuscript

---

Extended author information available on the last page of the article

Sherin & Van Es, 2009). They further need to collect, aggregate, and evaluate information about their students to make evidence-based decisions regarding which action to take to best guide their students (Heitzmann et al., 2019; Shavelson & Stern, 1981).

A promising approach for teachers to gain and train these necessary diagnostic reasoning skills is simulation-based learning (Chernikova et al., 2020b). In such learning environments, teachers engage in diagnostic activities while working on cases that simulate segments of reality (Heitzmann et al., 2019). Working on authentic cases not only provides opportunities to acquire and practice diagnostic reasoning (Hege et al., 2018) but also prepares students for situations they are likely to encounter in their professional future (Berman et al., 2016).

In this study, we used a simulation-based learning environment featuring cases to support teachers in gaining the strategic knowledge that is needed for diagnostic reasoning with respect to two common learning difficulties: attention deficit hyperactivity disorder (ADHD) and dyslexia. We compared a whole case format that presents all information about a student at once to a serial-cue case format that presents information about a student step by step. We investigated if participants perceive the serial-cue case format as more authentic and whether they felt more involved with serial-cue format cases. Further, we tested whether one format is more efficacious in promoting strategic knowledge and if this effect depends on prior conceptual knowledge.

Theory and prior evidence

Teachers' diagnostic reasoning

To judge whether their instruction is effective and to evaluate what their students need, teachers constantly observe and interpret student behavior and student work (Shavelson & Stern, 1981). Such diagnostic reasoning (Heitzmann et al., 2019) is a key competence to enact core components of teacher professionalism (Grossman, 2021). For example, elevating student thinking to the focal point of the instruction through responsive teaching practices (Robertson et al., 2016) or identifying behaviors that potentially prevent or delay children's academic achievement (Artelt & Gräsel, 2009; Hoth et al., 2016).

Diagnostic reasoning is a potentially cyclic process of noticing initial indicators, evaluating, hypothesizing, and collecting further evidence until enough information is available to make a decision (Heitzmann et al., 2019; Custers et al., 2000; Gruber, 2013). It involves conceptual knowledge—knowledge that helps teachers notice indicators of struggle, progress, or success. It also involves strategic knowledge—knowledge that helps teachers decide what action to take to gather more evidence (Förtsch et al., 2018; Gruber, 2013). Given this conceptualization, diagnostic reasoning is similar to professional vision. Professional "vision is not a purely mental process but instead is accomplished through the competent deployment of a complex of situated practices in a relevant setting" (Goodwin, 1994, p. 626). Teacher professional vision is separated into selective attention (where to prioritize attention in a busy classroom) and knowledge-based reasoning (what kind of teacher knowledge is used to reason about the noticed events) (Sherin & Van Es, 2009)—situation-specific skills that are much alike the skills we described as essential for diagnostic reasoning. The distinctive feature of diagnostic reasoning is that it is more of a problem-solving approach, involving a sequential process of engaging in epistemic activities to gather evidence (Fischer et al., 2014; Kramer et al., 2021) with a clear orientation towards actively seeking information (Förtsch et al., 2018). Thus, while the teacher is seen as an observer and decision-maker in

both professional vision and diagnostic reasoning conceptualizations, diagnostic reasoning highlights the teacher's role as an active investigator in professional situations (Kramer et al., 2021).

While teachers' professional practice will be characterized by on-the-fly-decision-making in busy classroom situations with many simultaneous goals and competing demands (Helleve et al., 2023; Tripp, 2011), teachers will also have to figure out why a student struggles with reading and writing but performs above average in math class. These situations are where investigation-related situation-specific skills are needed, skills that span from within the classroom (observing student behavior) to outside the classroom (investigation of prior records, homework, etc.). Yet, pre-service teachers get only few opportunities to practice diagnostic reasoning (Chernikova et al., 2020a; Heitzmann et al., 2019) and would benefit from opportunities that prepare them to apply diagnostic reasoning during problem-solving in professional practice (Heitzmann et al., 2019; Helleve et al., 2023). Further, having strategic knowledge and knowing how to actively gather information may also serve teachers in these complex situations where immediate decisions must be made.

The potential of simulation-based learning for teacher education

Great potential with respect to supporting pre-service teachers in learning to reason diagnostically is attributed to simulation-based learning (Chernikova et al., 2020a; Jossberger et al., 2022; Lehtinen, 2023; Sommerhoff et al., 2023). Simulations are environments that resemble reality or model real systems (de Jong, 2011). In simulations, students engage with authentic problems of practice—situations that are typical or to be expected in their future professional reality (e.g., Fischer et al., 2022). Simulations often use cases, for example, of a student or a patient (Bateman et al., 2013; Kiesewetter et al., 2020), who is introduced to the learner with text vignettes, videos, or else (Sykes & Bird, 1992).

Case-based learning (Kolodner, 1992) is a prominent approach in teacher education as it not only exemplifies real situations but also helps pre-service teachers to connect theory with practice and reflect on the interpretive variability of situations (Darling-Hammond & Hammerness, 2002; Helleve et al., 2023; Merseth & Lacey, 1993). Cases in the present study are referred to as *simulated cases* because the case design has some intentionalities that go beyond a "realistic narrative from classrooms and schools" (Helleve et al., 2023, p. 62). Simulated cases are designed to not only describe a real situation but also to create the impression of a genuine person within the described context. This approach is intended to evoke a more immersive and authentic experience for the participants. Further, simulated cases include various sources of evidence. This intentional inclusion of supplementary information aims to prompt analytical engagement with the case, encouraging participants to assess how the additional details influence their perceptions and whether they can affirm or challenge their beliefs about the case. With this design, simulated cases lend themselves particularly well for practicing diagnostic reasoning. As students study case descriptions, choose to gather more evidence, manipulate information, and add their own thoughts (Berman et al., 2016; Chi et al., 2018) to then use all the available information to make decisions with respect to the case (Heitzmann et al., 2019; Okuda et al., 2009), they "build rich mental representations" (Mamede et al., 2014, p. 121) of the problem. Students apply conceptual knowledge to case specifics to become more and more adept at interpreting, evaluating, and synthesizing evidence and making information-based decisions as to what next steps best to take (Thistlethwaite et al., 2012). Thus, as case-based learning foundationally "aims to cultivate analytic skills in the application of ideas and to convey theoretical knowledge

in a form useful to the interpretation of situations, the making of decisions, the choice of actions, and the formation of plans and designs" (Sykes & Bird, 1992, p. 469), it promotes strategic knowledge (Sykes & Bird, 1992; Thistlethwaite et al., 2012).

Simulated cases may be preferable to real-life practice as the complexity and unpredictability of real situations potentially overwhelm learners, especially in beginning stages of knowledge acquisition or training (Stegmann et al., 2012; Grossman et al., 2009). Another advantage of simulations is that learners can be guided to (repeatedly) practice specific parts of a task (van Merriënboer & Kirschner, 2017). A further benefit of simulations is that the costs of making mistakes are rather low in comparison to real life where accurate assessment of students' behavior, knowledge, and skills is crucial to provide them with appropriate and aptly suited support which impacts students' future learning, achievement, and school success.

Thus, simulations activate learners cognitively through engagement with authentic problems and create a version of reality that is manageable, and a save space, allowing for targeted learning and practice of core professional skills without risking unintended negative consequences for actual students.

Perceived authenticity and cognitive involvement

In this study, we designed cases with multiple intentions of authenticity: (a) "to emulate the work of professionals of a certain discipline," (b) "to reflect experiences from real/daily life," and (c) "to create personally meaningful learning activities" (Nachtigall et al., 2022, p. 1483). These design intentions for authenticity speak to a key element of simulations but also any other learning design that utilizes methods to enhance the transfer of learning, i.e., the application of acquired knowledge and skills in a new (real-world) setting. For example, constructivist and situated approaches to learning argue that this goal is only achieved by approximating or even equating the situation or context in which knowledge and skills are learned and applied (Brown et al., 1989; Collins et al., 1988). Authenticity hereby increases the practical or real-life relevance of the task or content which in turn positively influences learners' engagement, interest, and motivation. Meta-analytic evidence suggests that "simulations with an overall high authenticity do have greater effects than simulations with a lower authenticity" but that "even simulations with low authenticity still have large effect sizes, exceeding those of many other forms of instruction" (Chernikova et al., 2023, p. 523). Aligning with other evidence for "moderate to large effects of authentically designed learning settings on both cognitive and motivational learning outcomes (Nachtigall et al., 2022, p. 1506).

Yet, it is often a challenge to balance authenticity of simulations and the potentially associated cognitive demand (e.g., Blomberg et al., 2013). While high-fidelity simulations include tactile, auditory, and visual stimuli, not every simulation can include all these factors that make simulations feel real (Decker et al., 2008). Carefully aligning design with pedagogical or learning theory can make low-fidelity simulations that only use visual stimuli feel authentic (Grossman et al., 2014; Hamstra et al., 2014). This is because of the functional correspondence between simulations and reality—aligning the simulations' (functional) properties with the learning goals of the task (Hamstra et al., 2014)—might be more important than physical resemblance (Chernikova et al., 2023).

Even though a system, environment, or materials were intentionally designed to make them authentic (structurally and/or functionally), learners might not necessarily perceive them as such (e.g., Barab et al., 2000). The mechanism we see at work during

simulation-based learning aligns with Betz et al. (2016), who describe learners' *perception* of authenticity as key element of learning in environments that are designed to be authentic and aim to foster different motivational, affective, cognitive, or behavioral learning outcomes. For instance, Nachtigall et al. (2018) compared two instructional environments, one intentionally designed to be more authentic, and observed no evidence for a difference in learners' perceptions of authenticity between the two environments. However, perceived authenticity was positively correlated with situational interest in both instructional conditions. Thus, designing for authenticity may not result in its perception yet the suggested benefits of authenticity may only arise when learners subjectively perceive the environment as authentic. This, and because "previous research has not focused on the effects of authentically contextualized learning settings on learners' perceived authenticity" (Nachtigall et al., 2022, p. 1506) warrants the investigation of perceived authenticity in research on the authenticity of learning environments. It further seems beneficial to adopt a differentiated perspective on perceived authenticity with respect to functional versus physical similarities between the simulation and activities or situations of professional practice as well (Chernikova et al., 2023; Hamstra et al., 2014).

Thus, next to realizing authentic learning experiences, simulations need to capture a learner's sustained attention to help them construct a mental model of the situation (Schubert et al., 2001). In other words, a simulation should allow for "the subjective experience of being in one place or environment, even when one is physically situated in another" (Witmer & Singer, 1998, p. 225). This "being present" is the feeling of being fully involved in an experience (Vorderer et al., 2004). Such cognitive involvement may increase the learning that is happening while engaging with the simulation (e.g., Stevens & Kincaid, 2015). For example, Pickal et al. (2022) found that the perceived authenticity of a simulation did not predict the accuracy of diagnostic reasoning in a learning environment, however, the learners' involvement did.

In conclusion, authenticity is important but operates on multiple levels and might not be effectively supporting learning when the simulation not also engage learners cognitively. These points can guide the design of specific features of a simulation, such as the format of simulated cases.

The format of simulated cases and prior knowledge

Simulated cases can be presented in different formats which are mostly discussed in medical education, a field that already utilizes simulation-based learning frequently (Kiesewetter et al., 2020). The whole case format presents all information about the case upfront. That is, the entire case including all associated information is available from the minute the learner gets involved with the case (Al Rumayyan et al., 2018) remains available throughout the entire interaction with the case (Schmidt & Mamede, 2015). In professional practice, however, information is typically not just available but intentionally gathered (Heitzmann et al., 2019). For example, a teacher notices that a student persistently mixes up b's and d's. Based on this observation, the teacher pays more attention to other signs of dyslexia and eventually lets the student take a test. Mimicking this process of diagnostic reasoning, cases are often designed in a serial-cue format (Schmidt & Mamede, 2015). A serial-cue format case provides information to a learner gradually and only if the learner actively requests that information (e.g., Al Rumayyan et al., 2018). For example, the teacher who noticed a student to mix up b's and d's only learns whether this student also struggles in other subjects by asking their colleagues.

While the serial-cue format has a high face validity—giving learners a more authentic practice of how diagnosing plays out in real life—it possibly poses a challenge for learners with less prior knowledge in a domain (Schmidt & Mamede, 2015). The human cognitive system includes domain-unspecific functions that allow us to engage in goal-directed behavior and reasoning processes (e.g., Conway et al., 2002; Friedman & Miyake, 2017; Oberauer et al., 2003) as well as a "large storage of organized knowledge structures in long-term memory with effectively unlimited capacity and duration" (Kalyuga, 2007, p. 510). Existing knowledge representations in long-term memory help to validate and structure new incoming information (Dochy et al., 1999). Experts in any domain have many schemas—domain-specific knowledge structures and procedures—at their disposal which they utilize to execute routine tasks but can also draw on when solving new problems (e.g., Kalyuga, 2007).

In a learning context, learners with less prior knowledge typically benefit from instructional support that guides their reasoning by providing solution steps during problem-solving activities (Kalyuga, 2007; Sweller et al., 2019). These "external" knowledge structures compensate for the lack of existing internally stored structures that learners with more prior knowledge possess. A serial-cue case requires the learner to generate a hypothesis and make decisions as to what evidence to collect based on very little information, a task which learners are much more likely to succeed at when they have knowledge structures they can fall back on. Without a strong prior knowledge base, learners are likely resorting to weak problem-solving strategies (Newell & Simon, 1972; Sweller, 1988); deciding their next moves or generating their hypotheses based on superficial and irrelevant aspects of the problem (Atkinson et al., 2000; Kalyuga, 2007; Renkl, 2014). This resource-inefficient approach to solving the task may not help achieving the desired learning goals of acquiring domain knowledge (schemas) and gaining diagnostic reasoning skills that are applicable to future cases (Kalyuga, 2007) because the less prior knowledge a learner has, the more difficult it is for them to identify relevant information and connect new with existing information (Amadieu et al., 2009).

The whole case format would counteract such inefficient problem-solving approaches because its design acts as an external support system to compensate for missing knowledge structures. Learners are facing lower problem-solving demands when all the case information is presented upfront as they do not have to identify which option will potentially be a source of relevant information, review the data once they made choice what to look at, critically assess the evidence the source provides, compare against the problem and the other information gathered, synthesize the evidence from multiple sources and go through this process iteratively with every new evidence source that is selected. Instead, the learner can focus on working through the provided evidence, thereby gaining a knowledge base, and then concentrate on the evaluation of the presented evidence to discern the most relevant information for making a decision (Schmidt & Mamede, 2015).

Of course, as diagnostic reasoning is a complex task, there are still reasoning demands. Not all information given in the case description is equally important and learners still need to sort through the evidence as considering every piece of evidence just because it is available would not be an efficient nor goal-directed process. However, these demands are much reduced in comparison to the serial-cue case, which, on top of the many more processing requirements, also carry the risk of "premature closure" (Norman et al., 2017) for learners who do not have knowledge structures to fall back on. These learners might feel they have gathered enough evidence, see no benefit in requesting more information, and close the investigation, shutting down the diagnostic process too early. The risk lies here in disregarding relevant evidence,

focusing on only one piece of evidence, limiting one's perspective to a narrow set of hypotheses, or biasing confirming evidence (Custers et al., 2000).

A prior study that compared these two case formats did not find evidence that one format was more beneficial than the other or that the case format effect on (medical) students' diagnostic reasoning depends on prior knowledge (Kiesewetter et al., 2020). We are adding to the empirical evidence base by testing the comparative effectiveness of simulated case formats in teacher education.

The present study

A core aspect of teaching practice can be described as diagnosing students' progress and needs to find out in which ways to best support their learning (Grossman, 2021). Pre-service teachers can benefit from simulations, opportunities to engage in and practice diagnostic reasoning that prepare them for the demands of their professional future (Chernikova et al., 2020a, b). We designed an environment including simulated cases in the context of diagnostic reasoning to recognize, interpret, and take appropriate diagnostic steps toward an accurate diagnosis of two common learning difficulties: dyslexia and attention deficit hyperactivity disorder (ADHD), both of which are severe developmental disorders acknowledged by ICD-10 and DSM-V (World Health Organization, 2019; American Psychiatric Association, 2022). As teachers are more trained for diagnosing their students' motivation, interest, or understanding, recognizing indicators of learning difficulties such as ADHD or dyslexia can be challenging (e.g., Scahill & Schwab-Stone, 2000; Shaywitz et al., 2008). We chose dyslexia and ADHD as the content for our simulation to allow participants to gain and practice diagnostic reasoning on issues that are probably unfamiliar but relevant to their professional future.

We investigate effects of the serial-cue and whole case format on perceived authenticity, cognitive involvement, and strategic knowledge.

The serial-cue case format enables learners to make choices in their diagnostic process very similar to how they would act it out in real life. Utilizing this format might lead to high perceived authenticity and involvement, the proposed mechanism for making authentic learning scenarios effective for cognitive and motivational outcomes (Betz et al., 2016; Nachtigall et al., 2022). We hypothesize that learners in the serial-cue case format condition give higher authenticity ratings (H1) and higher involvement ratings (H2) than learners in the whole case format condition.

Given the evidence that tasks with lower problem-solving demands are usually more beneficial for learners with less prior knowledge (Kalyuga, 2007; Sweller et al., 2019), it might be worth to trade-off authenticity elements (serial-cue case format) for an instructional design that matches individual learners' needs (whole case format). We hypothesize that learners with less prior conceptual knowledge of ADHD and dyslexia gain more strategic knowledge of these learning disorders with the whole case and learners with more prior conceptual knowledge of ADHD and dyslexia gain more strategic knowledge of these learning disorders with the serial-cue case format (H3, disordinal interaction effect).

Method

Participants, prior knowledge training, and study design

A total of 118 pre-service teachers (86% women, 13% men, 1% non-binary) for primary school and higher track secondary school completed the study. We recruited pre-service teacher from all school types and from all semesters as courses on learning difficulties were elective and not tied to a specific semester, allowing them to be taken at any point during their studies. On average, participants were 23 years old ($SD = 4.10$; $min = 18$, $max = 40$) and in their 5th semester ($SD = 3.40$; $min = 1$, $max = 13$). Participants received 35€ compensation.

To ensure that prior conceptual knowledge of the targeted learning content varied sufficiently in our sample, we gave one group a short and the other group an extended input session about ADHD and dyslexia prior to the study. To keep the input sessions at the same length, participants in the short session were also given input on information processing models.

We used a between-subjects design with the independent variable case format (whole versus serial-cue), assessed prior knowledge (moderating variable) before, and strategic knowledge (dependent variable) after the simulation. We used stratified randomization to make sure that participants from the short and the long input sessions were approximately equally distributed across the two experimental conditions ($n_1 = 60$ serial-cue, $n_2 = 58$ whole).

Design for authentic learning

We categorize the design elements for creating an authentic learning environment according to Nachtigall et al. (2022). We (a) used technology; (b) designed complex cases; (c) utilized real-life materials or cultural tools for each simulated case; and (d) our participants went through an inquiry investigation to come to a diagnosis for each case. We outline these design elements for authenticity as we describe the details of our learning environment and the simulated cases.

Learning environment

We designed for authenticity using simulated cases as well as artifacts and materials borrowed from the cultural setting we are approximating (Radinsky et al., 2001). To create an inquiry investigation, we embedded our simulated cases and cultural artifacts in the computer-based simulation learning software CASUS (Kiesewetter et al., 2020; Fischer et al., 2005) (Fig. 1). For smooth navigation, we also administered the prior knowledge training, tests, ratings, etc. through the simulation in the CASUS environment.

Simulated cases

Participants engaged with eight simulated cases during the learning phase (four involved symptoms of ADHD and four involved symptoms of dyslexia, however, despite displaying symptoms, not all simulated students actually experienced these specific learning
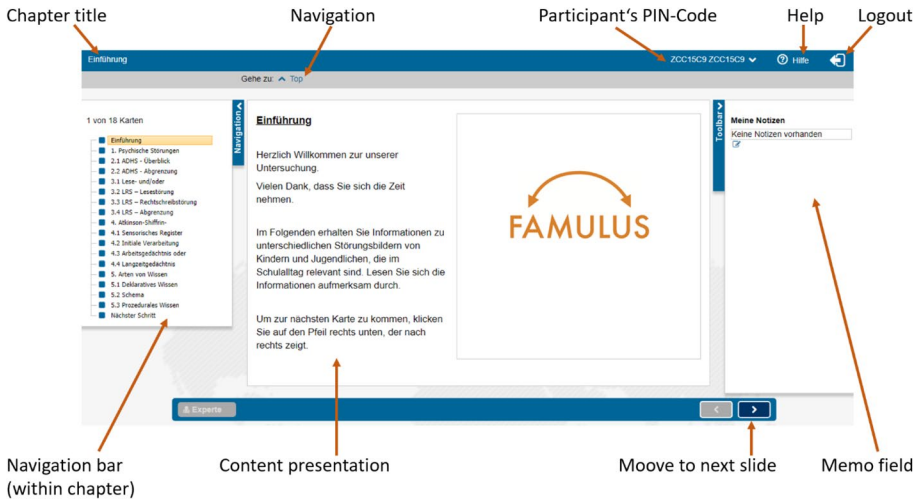
Fig. 1 Welcome page and navigation in the simulation-based environment CASUS

difficulties). Each case contained a description of the situation and a simulated student. For example, "You are a 4th grade elementary school teacher. The school year has just begun. Before the summer break, you noticed that Annika (9 years old) does not like to read aloud and completed the year with the grade "insufficient" in reading and writing." A vignette introduced the simulated student. The vignette described the student's social and study behavior and their attitude toward learning/at school. The simulated students in the vignettes spanned from 1st grade to 6th grade, capturing the transition between primary and secondary school to maximize relevance for pre-service teachers at all school tracks (see Stadler et al., 2021).

Additional information included worksheets, report cards, observations from the teacher in the classroom, protocols from parent-teacher conferences, conversations with other teachers, or a conversation with the student. Using such real-life materials and cultural tools is a key element for making the diagnostic process of gathering, synthesizing, and interpreting available evidence feel as authentic as possible.

Case format

In the whole case format condition, participants see all information about the simulated case at once as one long consecutive text. Participants scroll up and down to view the entire case narrative (Fig. 2).

In the serial-cue case format condition, participants view multiple buttons (e.g., "teacher observation," and "worksheets") to click on and are asked to make a diagnostic move. Once participants click on a button, the corresponding section from the simulated case narrative is shown. Participants can toggle back and forth between the intro page and pages with additional evidence and open as many evidence pages as are available (Fig. 3).

Fig. 2 Example simulated case: whole case format

Fig. 3 Example simulated case: serial-cue case format

Procedure

Participants completed the study in a lab (multiple people at the same time). They were shown how to navigate in CASUS. Casus creates unique and anonymous logins and logs all inputted data automatically. Participants were asked not to collaborate. They first completed the prior knowledge training, then the pretest, then the intervention. After

four simulated cases there was a break of 10–15 min after which participants continued with the remaining set of four cases and then completed the posttest.

Measures

Authenticity

Participants' perceived authenticity was assessed with three Likert-scale items after they completed the simulated cases: (1) I perceive the learning environment as authentic, (2) The learning environment felt like a real-life professional situation, and (3) The experience in the learning environment was similar to an experience in a real-life professional situation. These items were adapted from scales used in Seidel et al. (2011) and Schubert et al. (2001). The items were answered on a 5-point scale from 1=strongly disagree to 5=strongly agree. A mean score across all three items was used as indicator for perceived authenticity of the simulation. The scale showed high internal consistency: Cronbach's $\alpha=0.92$.

Involvement

Involvement was measured with four Likert-scale items after the simulation to assess participants' involvement with the cases and the simulated environment: (1) I was strongly focused on the situation, (2) I momentarily forgot that I was participating in a study, (3) I immersed myself mentally in the situation, and (4) I was fully concentrated on the situation. These items were adapted from Vorderer et al. (2004). The items were answered on a 5-point scale ranging from 1=strongly disagree to 5=strongly agree. The mean across all four items was calculated as an indicator of involvement and the scale showed acceptable reliability: Cronbach's $\alpha=0.67$.

Conceptual knowledge

Conceptual knowledge was measured at pretest to assess prior knowledge. Conceptual knowledge was assessed with 14 multiple choice items that each had four answer options and one correct answer. For example, "Which of the following is not one of the cardinal symptoms of ADHD?" with the answer options (a) Inattentiveness, (b) Hyperactivity, (c) Impulsivity, and (d) Impatience (correct option). For each correct answer, participants received one point and could thus achieve a minimum of 0 and a maximum of 14 points. The sum score was used as an indicator of conceptual knowledge. As the 14 items used to assess prior conceptual knowledge involve both questions about ADHD and dyslexia, we do not assume that the scale reflects a unidimensional construct and thus do not report internal consistency. Instead, we report the variance inflation factor (VIF) suggested for assessing the validity of formative constructs (Stadler et al., 2021; Taber, 2018). The VIF is an indicator of the redundancy in the scale. The VIFs ranged from 1.05 to 1.31 for the 14-item scale of conceptual knowledge, below the suggested cut-off of 3.3 for an acceptable degree of multicollinearity in the scale.

Strategic knowledge

Strategic knowledge was measured at posttest with four key feature cases (Page et al., 1995), two about ADHD and two about dyslexia. Each key feature case included a short case description of a few sentences describing a student's behavior and other observations or background information. The key feature cases differed from the simulated practice cases in length, detail about the simulated student, and response format. While participants added their diagnostic decision and reasoning for it into an open response text box during the intervention, posttest cases were followed by two questions with a choose-all-that-apply answer format.

The first question (diagnosis) asked participants to "select all diagnoses that appear most likely to be correct based on the available information. Please select N options". For each key feature case, it was specific how many options to choose, this number corresponded to the number of correct options for the question. Available options ranged from 7 to 10 across the four key feature cases. Each option could either be correct or incorrect. Correct options should be selected, incorrect options should not be selected. Participants received one point for correctly selecting and correctly not selecting answer options from the list.

The second question (strategy) for each key feature case asked participants "Which steps will you take to confirm or disconfirm the diagnosis of X?" where X was the correct diagnosis for each case (not the diagnoses, the participant selected as most likely correct in question (1). The answer options for question (2) ranged from 7–10 across the key feature cases. Each option was either correct or incorrect. Participants received one point for correctly selecting and correctly not selecting answer options from the list.

This scoring procedure resulted in a mean score for diagnosis and a mean score for strategy for each of the four key feature cases. The sum across these total eight mean scores was calculated to indicate the strategic knowledge of a participant. Thus, the final strategic knowledge could range from 0 to 8. We assume that strategic knowledge reflects a multidimensional construct (knowledge about two distinct learning disorders). The VIF analysis indicates a range of 1.03 to 1.10 for the strategic knowledge measure, demonstrating almost no collinearity between the items.

Control variable

Time on task was operationalized as the time participants took to complete the eight simulated cases, automatically logged by CASUS in seconds, transformed to minutes for easier interpretation, and log transformed to account for non-normal distribution (van der Linden, 2016). On average, participants spent 52 min (SD = about 16 min) on the eight simulated cases. Time on task was controlled for in the test of H3, the effect of case format and prior conceptual knowledge on acquisition of strategic knowledge.

Statistical analyses

We report one-tailed tests for directional hypotheses (manipulation check, hypothesis 1 and 2) and indicate this with $p_{one\text{-}tailed}$. We use a 5% alpha level to judge statistical significance. We do not apply Bonferroni correction even though we are conducting multiple tests using the same predictor variable (case format) because we make statistical

claims "for each individual test in the absence of an omnibus null hypothesis about which all of the tests speak collectively and for which the Type-1 error rate will be α by these corrections" (García-Pérez, 2023, p. 15). We used JASP (JASP Team, 2024) and the SPSS Macro PROCESS version 4.2_beta (Hayes, 2018) to estimate the moderated regression model to test H3. We estimated heteroscedasticity-consistent standard errors (HC3) and 95% bootstrapped confidence intervals (5000 samples) and reported unstandardized regression coefficients. We estimate Bayes factors to follow up when the frequentist approach results in inconclusive evidence ($p > 0.05$). Bayes factors quantify evidence for a given hypothesis and allow conclusions about which of two hypotheses is more likely. A Bayes factor of 3–10 is considered moderate and > 10 strong evidence for one hypothesis over another (Lee & Wagenmakers, 2014).

Results

Manipulation check and correlations

We used an independent samples t-test with training as predictor and prior conceptual knowledge as outcome. Prior conceptual knowledge was $M = 8.53$ ($SD = 1.70$) in the short input and $M = 9.19$ ($SD = 1.56$) in the long input group, a statistically significant difference $t(116) = 2.21$, $p_{one\text{-}tailed} = 0.015$, $d = 0.41$, 95% CI for Cohen's d $[-\infty, -0.10]$. We thus assume to have successfully increased variability of prior conceptual knowledge about ADHD and dyslexia in our sample.

Correlations are reported in Table 1. We controlled for involvement in the test for H1 and for perceived authenticity in the test for H2 because these variables were positively correlated.

Table 1 Correlations

		Prior conceptual knowledge	Perceived authenticity	Involvement	Time on task
Perceived authenticity	Pearson's r	0.028			
	p-value	0.767			
Involvement	Pearson's r	0.087	0.472		
	p-value	0.347	< 0.001		
Time on task	Pearson's r	0.126	0.122	− 0.064	
	p-value	0.174	0.186	0.494	
Strategic knowledge	Pearson's r	0.226	− 0.063	0.164	− 0.012
	p-value	0.014	0.495	0.076	0.894

$N = 118$ for all estimated correlations

Case format and perceived authenticity (H1)

To test whether participants perceived the serial-cue case format as more authentic than the whole case format simulation, we used a one-way ANCOVA with case format as the predictor, authenticity ratings as the outcome, and involvement ratings as the control variable. Results indicate that participants judged the authenticity almost the same in the serial-cue condition ($M = 3.4$, $SD = 1.17$) and the whole case ($M = 3.22$, $SD = 1.08$). This difference was statistically not significant $F(1115) = 0.04$, $p_{one\text{-}tailed} = 0.420$, $\eta_p^2 < 0.001$.

To learn more from this inconclusive result, we followed up with a Bayesian ANCOVA. Results indicate that the control variable involvement is the best predictor for authenticity. Our data are about 5 times more likely under the model including involvement as the only predictor (best model) than under the hypothesized model that includes involvement and case format as predictors ($BF_{01} = 5.1$). From this result we conclude that case format is not a key factor with respect to how authentic learners perceive the simulation. Instead, for perceiving the simulation as authentic, it seems more relevant if and to what degree learners feel able to concentrate on and immerse themselves into the situation.

Case format and involvement (H2)

We tested whether involvement was experienced to a higher degree in the serial-cue than the whole case format condition through a one-way ANCOVA with case format as the predictor, involvement ratings as outcome, and authenticity ratings as control variable.

Participants felt more involved in the serial-cue ($M = 3.9$, $SD = 0.6$) than in the whole case ($M = 3.63$, $SD = 0.76$) condition. This difference was statistically significant $F(1115) = 4.16$, $p_{one\text{-}tailed} = 0.022$, $\eta_p^2 = 0.04$. This means that students who worked on cases in the serial-cue format reported higher cognitive involvement than those who were in the whole case format condition.

Interaction of case format and prior conceptual knowledge (H3)

To test whether the effect of case format is dependent on prior conceptual knowledge, we estimated a moderated regression model including case format, prior conceptual knowledge, and the interaction of case format and prior conceptual knowledge as predictors of the outcome strategic knowledge (controlling for time on task).

The model was not statistically significant $F(4113) = 1.74$, $p = 0.146$, $R^2 = 0.06$. The effect of case format was not statistically significant $b = -0.29$, $p = 0.527$, CI_{boot} 95% $[-1.13, 0.58]$. The effect of prior conceptual knowledge was $b = 0.05$, $p = 0.127$, CI_{boot} 95% $[-0.01, 0.11]$. The interaction effect (case format*prior knowledge) was $b = 0.02$, $p = 0.625$, CI_{boot} 95% $[-0.07, 0.12]$. Simple slopes analysis indicated that the interaction of case format and prior knowledge was not statistically significant for levels of the moderator (Table 2).

Descriptively, this indicates that with increasing prior conceptual knowledge, the difference between the two case formats becomes smaller. However, neither the main effect of the case format nor the interaction case format*conceptual prior knowledge was statistically significant. Thus, we estimated a Bayesian moderated regression to quantify

Table 2 Conditional effects of case format on strategic knowledge at values of the moderator prior conceptual knowledge

Prior conceptual knowledge	Effect	p	95% LLCI	95% ULCI
Mean—1SD = 7.20	−0.11	0.350	−0.34	0.12
Mean = 8.86	−0.07	0.403	−0.23	0.09
Mean + 1 SD = 10.51	−0.03	0.813	−0.26	0.20

LLCI lower level confidence interval, *ULCI* upper level confidence interval

evidence for the null hypothesis relative to the alternative hypothesis (the effect of case format on posttest strategic knowledge depends on prior conceptual knowledge).

Results of the Bayesian analysis show that the best model is the model including only prior conceptual knowledge as a predictor for strategic knowledge. In fact, this model is 12.3 times more likely than the hypothesized model including prior conceptual knowledge, case format, and the prior conceptual knowledge*case format interaction (BF_{01} = 12.35). The best model, however, is only 3.7 times more likely than the model including prior conceptual knowledge and case format (BF_{01} = 3.66) (Table 3).

This means there is strong evidence that prior conceptual knowledge is the best predictor of strategic knowledge and strong evidence for prior conceptual knowledge as best predictor over an interaction of case format and prior conceptual knowledge. In contrast, there is only anecdotal evidence for the prior conceptual knowledge model over the prior conceptual knowledge and case format model. This means that we cannot be sure whether there is or is not an effect of case format on strategic knowledge, even after this follow-up analysis.

Table 3 Bayesian analysis of H3 interaction effect: model comparison

Models	P(M)	P(M|data)	BF_M	BF_{01}	Error %
Prior conceptual model	0.200	0.570	5.295	1.000	
Null model	0.200	0.182	0.892	3.125	0.002
Case format + prior conceptual knowledge	0.200	0.156	0.738	3.656	2.726
Case format + prior conceptual knowledge + case format*prior conceptual knowledge	0.200	0.046	0.194	12.345	1.273
Case format	0.200	0.046	0.193	12.369	0.029

Bayesian ANCOVA including posttest strategic knowledge as dependent variable, case format as fixed factor, prior conceptual knowledge as covariate, and case format*prior conceptual knowledge. Results show BF_{01} which indicates evidence for the null model over the alternative model, given the data. We display models in comparison to the best model. Thus, results can be interpreted as in favor of the best model (e.g., indicated by a BF_{01} = 3.656, the best model is 3.7 times more likely than the case format + prior conceptual model, given our data)

Discussion

Diagnosing patterns of student thinking and development is considered a core teaching practice (e.g., Grossman, 2021). Using simulation-based learning to engage pre-service teachers in diagnostic reasoning, we compared simulated cases with a whole case format to a serial-cue case format and tested the case formats' effects on perceived authenticity, involvement, and whether strategic knowledge acquisition depends on prior conceptual knowledge of ADHD and dyslexia.

Case format, perceived authenticity, and involvement (H1 and H2)

We did not find evidence that learners perceived the serial-cue case format as more authentic than the whole case format. However, we found that learners in the serial-cue case format condition reported higher levels of being cognitively involved in the simulation and a positive correlation between perceived authenticity and involvement.

We propose to distinguish between two levels on which our simulation's authenticity operates (Chernikova et al., 2023; Hamstra et al., 2014): the authenticity of the simulated case (physical authenticity) and the authenticity of the diagnostic process (functional authenticity). We used the same case materials in cases with different formats. Thus, only the functional authenticity was varied. Yet, the difference in authenticity that we intended and expected, was not reflected in participants' ratings—mirroring how an intended authenticity manipulation did not result in perceived differences in Nachtigall et al. (2018). This might be because the diagnostic reasoning process was unfamiliar to pre-service teachers or because their idea of diagnosing in professional situations differed from the activities in the study. Furthermore, the pre-service teachers involved in our study may have had varying degrees of prior teaching experience in schools and at different schools. This experience could be an additional factor influencing how authentic the participants perceived the simulations with either of the two case formats. That is, those with more teaching experience may have been better positioned to judge the authenticity of the simulation.

Participants in the serial-cue case condition reported higher cognitive involvement than those in the whole case condition. We assume learners felt this higher involvement because serial-cue case format allowed them to choose how to approach the case and because they had to weigh whether the evidence was relevant or negligible. Learners in the whole case condition may have felt less cognitively involved because as they reviewed the case information, they may have felt more like reading a long text than actively making choices. As Pickal et al. (2022) showed, cognitive involvement is crucial for accurate diagnosing in simulation-based learning, the serial-cue case format should be preferred over the whole case format based on our results.

Case format and prior conceptual knowledge (H3)

Our study leaves us with inconclusive results with respect to whether one case format was more effective in supporting learners to gain strategic knowledge needed for diagnostic reasoning. We did find that there is likely no interaction between case format and prior conceptual knowledge.

While studies with "authentic learning materials" and "resemble real-life experiences" as goals for authenticity often showed positive effects on *motivational* outcomes, they often reported negative effects on *cognitive* outcomes (Nachtigall et al., 2022). Our study delivers

inconclusive evidence with respect to cognitive learning outcomes, aligning with this prior work and a prior study, that found no differences between the whole and serial-cue case format in medical education (Kiesewetter et al., 2020). Nachtigall et al. (2022) conclude that there are seemingly no clear-cut design elements or intentions of authenticity that promise robust positive effects on cognitive outcomes. We argue that the leverage point for fostering cognitive learning outcomes is neither the design elements themselves nor the intentions that went into designing authentic learning environments, but rather the learning processes that are elicited through design features. While authenticity is important to elicit learners' (initial) interest (or buy-in; Hamstra et al., 2014), what ultimately supports learners in abstracting a mental model might be the cognitive involvement as it means learners focus on and engage in learning relevant processes (Chi et al., 2018; Schubert et al., 2001).

It is thus especially puzzling that students in the serial-cue case condition reported higher cognitive involvement but, in contrast to Pickal et al. (2022), did not gain more strategic knowledge. The case format differences may have been too subtle to cause noticeable differences in strategic knowledge. Especially since the cases were embedded in an else very rich simulated environment (Kiesewetter et al., 2020). This may also explain why fewer or less detailed existing knowledge structures (Kalyuga, 2007) did not disadvantage learners in the serial-cue case condition. The entire learning experience may have been a novel and complex situation resulting in high processing demands in both case formats (Chernikova et al., 2020b). Posttest cases further only asked participants to choose a likely diagnosis from a list and then select the most promising next steps. The learning experience in its entirety may have prepared participants for this test. It is possible that differences between the case formats might occur if the diagnostic reasoning process was tested in more nuance. Thus, differences in acquired or improved knowledge structures—schemas—for the diagnostic process may simply gone undetected. Although there is robustness theory and we identified reasons why our study may not have detected an effect of case format on strategic knowledge, we are also contemplating the possibility of a null effect.

Moving authentic learning forward

To move authentic learning forward, we are synthesizing across our three research questions and associated results to generate theoretical assumptions, suggest future lines of research and to derive a design implication.

Theoretical implications

We propose two theoretical assumptions that can be tested in future research. First, we are wondering if there is an authenticity threshold. Our learners differed with respect to involvement, but not perceived authenticity although involvement and perceived authenticity correlated positively. We think this could suggest, that physical authenticity is what onboards learners to the simulation. This means learners need to perceive a simulation as authentic enough to take it seriously. Once a simulation is perceived as authentic (enough), there is no benefit of trying to achieve even more physical authenticity—it might even be counterproductive (Blomberg et al., 2013). Arguably, it may also be that enough functional authenticity results in this "buy-in" (Hamstra et al., 2014, p. 388). Once learners are on board, the simulated activity needs to sustain involvement, that is, foster effective cognitive processes to positively influence the target learning goal (Hamstra et al., 2014). Exploring the threshold idea could involve examining the relationship between physical and

functional authenticity. This investigation may uncover whether certain learning settings or goals benefit more from one type or a combination of both for achieving the required level of authenticity for effective learner engagement (onboarding).

Second, there may be dependencies, maybe even a causal mechanism, between case format, involvement, perceived authenticity, and learning. If authenticity onboards learners, cognitive involvement then elicits key cognitive processes that increase the likelihood of learning, it is possible that authenticity is a moderator ("only if I find the simulation authentic…") and involvement a mediator (explaining case format effects on learning outcomes) in the relationship of case format (or any other feature of the simulation) and cognitive learning outcomes. Another potential (causal) relationship might be that the serial-cue case format allows learners to immerse themselves more into the simulated situation which then as a result makes the experience more authentic in the learner's perception. Hence, functional authenticity may impact immersion, immersion goes hand in hand with cognitive involvement which then results in learning gains.

Implications for research

Based on our results, other absence of evidence for a case format effect on knowledge acquisition (Kiesewetter et al., 2020), evidence that authentic learning environments may not consistently enhance cognitive outcomes (Nachtigall et al., 2022), and the rationale why simulations are a meaningful learning approach (e.g., Heitzmann et al., 2019; Ledger et al., 2022), we propose redirecting research attention from design to exploring how to maximize effects of learning with simulated (serial-cue) cases for learners with varying prerequisites.

First, we suggest focusing on guidance during learning with simulated cases. This might be particularly helpful for learners with less prior knowledge who seem to struggle navigating the complex process of diagnostic reasoning in simulations (Kiesewetter et al., 2020). For example, through adaptive feedback that helps learners distinguish between irrelevant and relevant evidence or next steps in approaching a case (Sailer et al., 2023), with worked examples or through self-explaining during the diagnostic reasoning process (Bichler et al., 2022), or other guidance techniques (Sommerhoff et al., 2023).

Second, we suggest focusing on practice of specific aspects of diagnostic reasoning—for which simulations lend themselves particularly well (van Merriënboer & Kirschner, 2017). For example, learners could systematically practice evaluating evidence, synthesizing multiple evidence sources to avoid premature or biased diagnoses, or distinguishing relevant and irrelevant evidence.

Practical implications

Based on our results, we suggest using the serial-cue case format in the future. Even though it was not perceived as more authentic, it is advantageous in representing a practice situation of diagnostic reasoning (Schmidt & Mamede, 2015). It is not costly to design cases in the serial cue format and it has the positive effect of learners feeling cognitively involved. The resemblance of diagnosing in a serial-cue case format simulation to a real-life practice situation, and the learners' engagement with the materials, increases the likelihood that students transfer the practiced reasoning processes to similar situations in their professional future (Chernikova et al., 2020b).

Limitations

Our results could be strengthened by a trace data analysis (Fan et al., 2023) that could confirm whether students in the serial-cue case format condition worked through the information selectively and consecutively. It is possible that some learners in the serial-cue case format condition clicked on all available information buttons and created themselves a whole case format. However, they still made decisions as to what information to consult and in which order—a key feature of the serial-cue case format.

Our study does not shed light on how learners move through the serial-cue case format, which evidence sources are often or always consulted, and which are often or always neglected. This would have indications for case design and guidance as learners potentially neglect relevant evidence or place too much emphasis on evidence that is not relevant for a given case. Similarly, we also do not have insight into the cognitive processing or steps those learners in the whole case condition took. As there is no log data associated with the whole case format, think aloud protocols or interviews could be utilized to gain this data.

Another potential threat to perceived authenticity might be the application context in which our simulated case format and diagnostic reasoning investigation was situated. Teachers are certainly amongst the first persons to notice signs of learning difficulties but are not positioned, in fact, not permitted to make any formal diagnosis of ADHD or dyslexia (in Germany). Diagnosing in a teaching context rather entails finding the most effective teaching and support strategies in heterogenous student bodies to tailor instruction to students with varying aptitudes for learning. Therefore, pre-service teachers might not have perceived the tasks as authentic, as they did not consider diagnostic reasoning related to ADHD and dyslexia within the scope of their professional responsibilities. It remains uncertain whether participants were aware that the targeted learning outcome was the diagnostic process and diagnostic reasoning processes and that these processes are transferable to other contexts.

Conclusion

We conclude that cognitive involvement with the task is more important for learning outcomes than different case designs, designs for and perception of authenticity. Thus, we suggest using serial-cue cases in simulation-based teacher education, or other designs with functional authenticity to prioritize the learning goals and learning processes over the learning material design. Further, we conclude that future research into authenticity thresholds as well as into the relationship of authenticity, involvement, and learning will benefit the field. Yet, we propose that the most impactful future avenue is to choose a design for which there is a robust theoretical rationale or that has functional authenticity for a learning goal and then investigate how the effectiveness of this design can be maximized through feedback, adaptivity, or other forms of guidance that facilitate the targeted learning outcome through relevant cognitive processes.

Data availability Anonymized study data, analysis script, and materials are shared on the Open Science Framework (OSF) https://osf.io/4gr8t/.

Declarations

Ethics approval The studies involving human participants were reviewed and approved by the Ethics Committee of the Medical Faculty of LMU Munich (no. 17-249). The participants provided their written informed consent to participate in this study.

Competing interests The authors declare no competing interests.

References

Al Rumayyan, A., Ahmed, N., Al Subait, R., Al Ghamdi, G., Mohammed Mahzari, M., Awad Mohamed, T., Rotgans, J. I., Donmez, M., Mamede, S., & Schmidt, H. G. (2018). Teaching clinical reasoning through hypothetico-deduction is (slightly) better than self-explanation in tutorial groups: An experimental study. *Perspectives on Medical Education, 7*, 93–99. https://doi.org/10.1007/s40037-018-0409-x

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction, 19*(5), 376–386. https://doi.org/10.1016/j.learninstruc.2009.02.005

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev). https://doi.org/10.1176/appi.books.9780890425787

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

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214. https://doi.org/10.3102/00346543070002181

Barab, S. A., Squire, K. D., & Dueber, W. (2000). A co-evolutionary model for supporting the emergence of authenticity. *Educational Technology Research and Development, 48*(2), 37–62. https://doi.org/10.1007/BF02313400

Bateman, J., Allen, M., Samani, D., Kidd, J., & Davies, D. (2013). Virtual patient design: Exploring what works and why. A grounded theory study. *Medical Education, 47*(6), 595–606. https://doi.org/10.1111/medu.12151

Berman, N. B., Durning, S. J., Fischer, M. R., Huwendiek, S., & Triola, M. M. (2016). The role for virtual patients in the future of medical education. *Academic Medicine, 91*(9), 1217–1222. https://doi.org/10.1097/ACM.0000000000001146

Betz, A., Flake, S., Mierwald, M., & Vanderbeke, M. (2016). Modelling authenticity in teaching and learning contexts. A contribution to theory development and empirical investigation of the construct. . In C. K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016* (Vol. 2, pp. 815–818). International Society of the Learning Sciences.

Bichler, S., Stadler, M., Bühner, M., Greiff, S., & Fischer, F. (2022). Learning to solve ill-defined statistics problems: Does self-explanation quality mediate the worked example effect? *Instructional Science, 50*, 335–359. https://doi.org/10.1007/s11251-022-09579-4

Blomberg, G., Renkl, A., GamoranSherin, M., Borko, H., & Seidel, T. (2013). Five research-based heuristics for using video in pre-service teacher education. *Journal for Educational Research Online, 5*(1), 90–114. https://doi.org/10.25656/01:8021

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

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

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

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

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

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children, 8*(1), 2–10.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Custers, E. J., Stuyt, P. M., & Robbé, P. F. D. V. (2000). Clinical problem analysis (CPA): A systematic approach to teaching complex medical problem solving. *Academic Medicine, 75*(3), 291–297.

Darling-Hammond, L., & Hammerness, K. (2002). Toward a pedagogy of cases in teacher education. *Teaching Education, 13*(2), 125–135. https://doi.org/10.1080/1047621022000007549

de Jong, T. (2011). Instruction based on computer simulations. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 446–466). Routledge.

Decker, S., Sportsman, S., Puetz, L., & Billings, L. (2008). The evolution of simulation and its contribution to competency. *The Journal of Continuing Education in Nursing, 39*(2), 74–80. https://doi.org/10.3928/00220124-20080201-06

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*(2), 145–186. https://doi.org/10.3102/00346543069002145

Fan, Y., Rakovic, M., van Der Graaf, J., Lim, L., Singh, S., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2023). Towards a fuller picture: Triangulation and integration of the measurement of self-regulated learning based on trace and think aloud data. *Journal of Computer Assisted Learning, 39*(4), 1039–1395. https://doi.org/10.1111/jcal.12801

Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F., & Jünger, J. (2005). A modified electronic key feature examination for undergraduate medical students: Validation threats and opportunities. *Medical Teacher, 27*(5), 450–455. https://doi.org/10.1080/01421590500078471

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., & Fischer, M. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 2*(3), 28–45.

Fischer, F., Bauer, E., Seidel, T., Schmidmaier, R., Radkowitsch, A., Neuhaus, B. J., Hofer, S. I., Sommerhoff, D., Ufer, S., & Kuhn, J. (2022). Representational scaffolding in digital simulations–learning professional practices in higher education. *Information and Learning Sciences, 123*(11/12), 645–665. https://doi.org/10.1108/ILS-06-2022-0076

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

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204.

García-Pérez, M. A. (2023). Use and misuse of corrections for multiple testing. *Methods in Psychology, 8*, 100120. https://doi.org/10.1016/j.metip.2023.100120

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Grossman, P. (2021). *Teaching core practices in teacher education*. Harvard Education Press.

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. https://doi.org/10.1177/016146810911100

Grossman, R., Heyne, K., & Salas, E. (2014). Game-and simulation-Based approaches to training. In K. Kraiger, J. Passmore, N. R. dos Santos, & S. Malvezzi (Eds.), *The Wiley Blackwell handbook of the*

psychology of training, development, and performance improvement (pp. 205–223). https://doi.org/10.1002/9781118736982.ch12

Gruber, T. R. (2013). *The acquisition of strategic knowledge*. Elsevier.

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

Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.). Guilford Publications.

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. https://doi.org/10.1371/journal.pone.0204900

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. https://doi.org/10.14786/flr.v7i4.384

Helleve, I., Eide, L., & Ulvik, M. (2023). Case-based teacher education preparing for diagnostic judgement. *European Journal of Teacher Education, 46*(1), 50–66. https://doi.org/10.1080/02619768.2021.1900112

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

JASP Team. (2024). JASP (Version 0.18.3) [Computer software].

Jossberger, H., Breckwoldt, J., & Gruber, H. (2022). Promoting expertise through simulation (PETS): A conceptual framework. *Learning and Instruction, 82*, 101686. https://doi.org/10.1016/j.learninstruc.2022.101686

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kiesewetter, J., Sailer, M., Jung, V. M., Schönberger, R., Bauer, E., Zottmann, J. M., Hege, I., Zimmermann, H., Fischer, F., & Fischer, M. R. (2020). Learning clinical reasoning: How virtual patient case format and prior knowledge interact. *BMC Medical Education, 20*, 1–10. https://doi.org/10.1186/s12909-020-1987-y

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

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

Ledger, S., Burgess, M., Rappa, N., Power, B., Wong, K. W., Teo, T., & Hilliard, B. (2022). Simulation platforms in initial teacher education: Past practice informing future potentiality. *Computers & Education, 178*, 104385. https://doi.org/10.1016/j.compedu.2021.104385

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lehtinen, E. (2023). Can simulations help higher education in training professional skills? *Learning and Instruction, 86*, 101772. https://doi.org/10.1016/j.learninstruc.2023.101772

Mamede, S., Van Gog, T., Sampaio, A. M., De Faria, R. M. D., Maria, J. P., & Schmidt, H. G. (2014). How can students' diagnostic competence benefit most from practice with clinical cases? The effects of structured reflection on future diagnosis of the same and novel diseases. *Academic Medicine, 89*(1), 121–127. https://doi.org/10.1097/ACM.0000000000000076

Merseth, K. K., & Lacey, C. A. (1993). Weaving stronger fabric: The pedagogical promise of hypermedia and case methods in teacher education. *Teaching and Teacher Education, 9*(3), 283–299. https://doi.org/10.1016/0742-051X(93)90044-H

Nachtigall, V., Rummel, N., & Serova, K. (2018). Authentic does not equal authentic—how students evaluate the authenticity of learning activities in an out-of-school lab. *Unterrichtswissenschaft, 46*, 299–319. https://doi.org/10.1007/s42010-018-0020-1

Nachtigall, V., Shaffer, D. W., & Rummel, N. (2022). Stirring a secret sauce: A literature review on the conditions and effects of authentic learning. *Educational Psychology Review, 34*(3), 1479–1516. https://doi.org/10.1007/s10648-022-09676-3

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.

Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The causes of errors in clinical reasoning: Cognitive biases, knowledge deficits, and dual process thinking. *Academic Medicine, 92*(1), 23–30. https://doi.org/10.1097/ACM.0000000000001421

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*(2), 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

Okuda, Y., Bryson, E. O., DeMaria, S., Jr., Jacobson, L., Quinones, J., Shen, B., & Levine, A. I. (2009). The utility of simulation in medical education: What is the evidence? *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine: A Journal of Translational and Personalized Medicine, 76*(4), 330–343. https://doi.org/10.1002/msj.20127

Page, G., Bordage, G., & Allen, T. (1995). Developing key-feature problems and examinations to assess clinical decision-making skills. *Academic Medicine, 70*(3), 194–201.

Pickal, A. J., Engelmann, K., Girwidz, R., Neuhaus, B. J., & Wecker, C. (2022). Using simulations to foster pre-service teachers' diagnostic competences: What aspect of authenticity matters. In C. Chinn, E. Tan, C. Chan, & Y. Kali (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences - ICLS 2022* (pp. 1353–1356). International Society of the Learning Sciences.

Radinsky, J., Bouillion, L., Lento, E. M., & Gomez, L. M. (2001). Mutual benefit partnership: A curricular design for authenticity. *Journal of Curriculum Studies, 33*(4), 405–430. https://doi.org/10.1080/00220270118862

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*(1), 1–37. https://doi.org/10.1111/cogs.12086

Robertson, A. D., Atkins, L. J., Levin, D. M., & Richards, J. (2016). What is responsive teaching? In A. D. Robertson, R. Scherr, & D. Hammer (Eds.), *Responsive Teaching in Science and Mathematics* (pp. 1–35). Routledge.

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learning and Instruction, 83*, 101620. https://doi.org/10.1016/j.learninstruc.2022.101620

Scahill, L., & Schwab-Stone, M. (2000). Epidemiology of ADHD in school-age children. *Child and Adolescent Psychiatric Clinics of North America, 9*(3), 541–555. https://doi.org/10.1016/S1056-4993(18)30106-8

Schmidt, H. G., & Mamede, S. (2015). How to improve the teaching of clinical reasoning: A narrative review and a proposal. *Medical Education, 49*(10), 961–973. https://doi.org/10.1111/medu.12775

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

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education, 27*(2), 259–267. https://doi.org/10.1016/j.tate.2010.08.009

Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 51*(4), 455–498. https://doi.org/10.3102/00346543051004455

Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology, 59*, 451–475. https://doi.org/10.1146/annurev.psych.59.103006.093633

Sherin, M. G., & Van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*(1), 20–37. https://doi.org/10.1177/0022487108328

Sommerhoff, D., Codreanu, E., Nickl, M., Ufer, S., & Seidel, T. (2023). Pre-service teachers' learning of diagnostic skills in a video-based simulation: Effects of conceptual vs. interconnecting prompts on judgment accuracy and the diagnostic process. *Learning and Instruction, 83*, 101689. https://doi.org/10.1016/j.learninstruc.2022.101689

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. https://doi.org/10.1016/j.newideapsych.2020.100832

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

Stevens, J. A., & Kincaid, J. P. (2015). The relationship between presence and performance in virtual simu-
lation training. *Open Journal of Modelling and Simulation, 3*(02), 41–48. https://doi.org/10.4236/
ojmsi.2015.32005

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2),
257–285. https://doi.org/10.1016/0364-0213(88)90023-7

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

Sykes, G., & Bird, T. (1992). Chapter 10: Teacher education and the case idea. *Review of Research in Edu-
cation*, *18*(1), 457–521. https://doi.org/10.3102/0091732X018001457

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments
in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/
s11165-016-9602-2

Thistlethwaite, J. E., Davies, D., Ekeocha, S., Kidd, J. M., MacDougall, C., Matthews, P., Purkis, J., & Clay,
D. (2012). The effectiveness of case-based learning in health professional education. A BEME system-
atic review: BEME Guide No. 23. *Medical Teacher, 34*(6), e421–e444. https://doi.org/10.3109/01421
59X.2012.680939

Tripp, D. (2011). *Critical incidents in teaching* (classic). Routledge.

van der Linden, W. J. (2016). Lognormal response-time model. In W. J. van der Linden (Ed.), Handbook *of
Item Response Theory* (Vol. 1, pp. 289–310). Chapman and Hall/CRC.

van Merriënboer, J. J., & Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to
four-component instructional design* (3 ed.). Routledge.

Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, F., Böcking, S., Schramm, H., Gysbers,
A., Hartmann, T., Klimmt, C., Laarni, J., Ravaja, N., Sacau, A., Baumgartner, T., & Jäncke, P. (2004).
*MEC Spatial Presence Questionnaire (MEC-SPQ): Short Documentation and In structions for Appli-
cation*. https://www.researchgate.net/publication/318531435_MEC_spatial_presence_questionnaire_
MEC-SPQ_Short_documentation_and_instructions_for_application. Accessed 1 July 2024.

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

World Health Organization. (2019). *World health statistics overview 2019: Monitoring health for the SDGs,
sustainable development goals*. World Health Organization.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and
institutional affiliations.

Sarah Bichler. Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13,
80802 Munich, Germany. sarah.bichler@psy.lmu.de

Current themes of research

Learning analytics. individual learner differences. adaptive and responsive instruction. AI in education.

Most relevant publications in the field of Psychology of Education

Bichler, S., Stadler, M., Bühner, M., Greiff, S., & Fischer, F. (2022). Learning to solve ill-defined statistics
problems: does self-explanation quality mediate the worked example effect? *Instructional Science, 50*,
335–359. https://doi.org/10.1007/s11251-022-09579-4

Michael Sailer. Learning Analytics and Educational Data Mining, University of Augsburg, Universitätsstraße
10, 86159 Augsburg, Germany. michael.sailer@uni-a-de

Current themes of research

AI in education. simulation-based learning. learning analytics.

Most relevant publications in the field of Psychology of Education

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learning and Instruction, 83*, 101620. https://doi.org/10.1016/j.learninstruc.2022.101620

Elisabeth Bauer. Learning Analytics and Educational Data Mining, University of Augsburg, Universitätsstraße 10, 86159 Augsburg, Germany. elisabeth.bauer@uni-a.de

Current themes of research

Simulation-based learning. diagnostic reasoning. learning analytics. AI in education.

Most relevant publications in the field of Psychology of Education

Fischer, F., Bauer, E., Seidel, T., Schmidmaier, R., Radkowitsch, A., Neuhaus, B. J., … & Fischer, M. R. (2022). Representational scaffolding in digital simulations–learning professional practices in higher education. *Information and Learning* Sciences, *123*(11/12), 645–665. https://doi.org/10.1108/ILS-06-2022-0076

Jan Kiesewetter. Institute of Medical Education, LMU University Hospital, LMU Munich, Pettenkoferstr. 8a, 80336 Munich, Germany. jan.kiesewetter@med.uni-muenchen.de

Current themes of research

Diagnostic reasoning. simulation-based learning. resilience training.

Most relevant publications in the field of Psychology of Education

Kiesewetter, J., Sailer, M., Jung, V., Schönberger, R., Bauer, E., Zottmann, J., M., Hege, I., Fischer, F., & Fischer, M. R. (2020). Learning clinical reasoning: how virtual patient case format and prior knowledge interact. *BMC Medical Education, 20*(73). https://doi.org/10.1186/s12909-020-1987-y

Hanna Härtl. Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 Munich. hanna.haertl@t-online.de

Current themes of research

Diagnostic reasoning. simulation-based learning.

Most relevant publications in the field of Psychology of Education

N/A.

Martin R. Fischer. Institute of Medical Education, LMU University Hospital, LMU Munich, Pettenkoferstr. 8a, 80336 Munich, Germany. Martin.Fischer@med.uni-muenchen.de

Current themes of research

Diagnostic reasoning. simulation-based learning. faculty and curriculum development.

Most relevant publications in the field of Psychology of Education

Fink, M.C., Heitzmann, N., Reitmeier, V., Siebeck, M., Fischer, F., & Fischer, M.R. (2023) Diagnosing virtual patients: the interplay between knowledge and diagnostic activities. *Advances in Health Sciences Education*. https://doi.org/10.1007/s10459-023-10211-4

Frank Fischer. Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 Munich. Frank.Fischer@psy.lmu.de

Current themes of research

Diagnostic reasoning. simulation-based learning.

Most relevant publications in the field of Psychology of Education

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

Authors and Affiliations

Sarah Bichler[1,2] · **Michael Sailer[3]** · **Elisabeth Bauer[3]** · **Jan Kiesewetter[4]** · **Hanna Härtl[2]** · **Martin R. Fischer[4]** · **Frank Fischer[2]**

✉ Michael Sailer
    michael.sailer@uni-a.de

    Sarah Bichler
    Sarah.Bichler@psy.lmu.de

    Elisabeth Bauer
    elisabeth.bauer@uni-a.de

    Jan Kiesewetter
    jan.kiesewetter@med.uni-muenchen.de

    Hanna Härtl
    hanna.haertl@t-online.de

    Martin R. Fischer
    Martin.Fischer@med.uni-muenchen.de

    Frank Fischer
    Frank.Fischer@psy.lmu.de

[1] School of Education, University of California, Berkeley, CA, USA

[2] Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 Munich, Germany

[3] Learning Analytics and Educational Data Mining, University of Augsburg, Universitätsstraße 10, 86159 Augsburg, Germany

[4] Institute of Medical Education, LMU University Hospital, Ludwig-Maximilians-Universität München, Pettenkoferstr. 8a, 80336 Munich, Germany