Simulation-Based Learning in Higher Education: A Meta-Analysis

Olga Chernikova, Nicole Heitzmann, and Matthias Stadler
Ludwig-Maximilians-Universität in Munich

Doris Holzberger iD
TUM School of Education, Centre for International Student Assessment, Technical University of Munich

Tina Seidel iD
TUM School of Education, Technical University of Munich

Frank Fischer
Ludwig-Maximilians-Universität in Munich

Simulation-based learning offers a wide range of opportunities to practice complex skills in higher education and to implement different types of scaffolding to facilitate effective learning. This meta-analysis includes 145 empirical studies and investigates the effectiveness of different scaffolding types and technology in simulation-based learning environments to facilitate complex skills. The simulations had a large positive overall effect: g = 0.85, SE = 0.08; CIs [0.69, 1.02]. Technology use and scaffolding had positive effects on learning. Learners with high prior knowledge benefited more from reflection phases; learners with low prior knowledge learned better when supported by examples. Findings were robust across different higher education domains (e.g., medical and teacher education, management). We conclude that (1) simulations are among the most effective means to facilitate learning of complex skills across domains and (2) different scaffolding types can facilitate simulation-based learning during different phases of the development of knowledge and skills.

Knowledge application in more or less realistic situations has been shown to be important for the development of complex skills (e.g., Kolodner, 1992). Expertise development theories (e.g., Van Lehn, 1996) suggest that learners acquire high

levels of expertise in complex problem-solving tasks if they dispose of sufficient prior knowledge and engage in a large amount of practice. Practice opportunities ideally include authentic problems related to a professional field (e.g., Barab et al., 2000). However, in higher and further education programs, the opportunity to engage in real-life problem solving is limited. In addition, practice in real-life situations without systematic guidance can be overtaxing for students and come with risks and ethical issues—for example, when working with real students or patients without being systematically prepared. Moreover, real-life situations do not always provide enough practice opportunities as, for example, critical situations appear less frequently or require a lot of time before decisions lead to observable consequences. These limitations make practice in real-life situations a somewhat inaccessible and sometimes suboptimal learning space, particularly for novice learners. Therefore, approximations of practice in which the complexity is reduced (Grossman et al., 2009) can help engage learners in specific aspects of professional practice and are promising in order to avoid confusion and efficiently use resources for learning and instruction. These approximations of practice can be realized in higher education with simulations, which allow students to use authentic problems and also to create a learning environment to practice and facilitate the acquisition of target complex skills (e.g., Cook, 2014).

Simulations are increasingly often used in higher education settings. In STEM (science, technology, engineering, and mathematics) education (e.g., D'Angelo et al., 2014; Wu & Anderson, 2015), they are used to facilitate a deeper understanding of concepts and relationships between them, advance inquiry, problem solving, and decision making. A lot of research has been done in the field of medical education (Cook, 2014; Cook et al., 2013; Hegland et al., 2017), where simulations are used to advance diagnostic competences and motor and technical skills of prospective doctors, nurses, and emergency teams. Simulation-based learning also occurs in other fields, such as teacher education, engineering, and management (e.g., Alfred & Chung, 2011; Brubacher et al., 2015). The present meta-analysis focuses on higher education and, more specifically, on fields that strongly rely on interaction with other people at different levels (physical, cognitive, social, etc.)—for example, medicine, nursing, psychological counseling, management, teacher education, particular areas of engineering, and economics. Regarding this area of interest, little is known about for whom simulations are particularly helpful, what scenarios are effective, and what additional instructional support makes them effective for learners with different learning prerequisites. Synthesized results on the role of different features of simulations (e.g., duration, technology use) and instructional support (e.g., scaffolding) are lacking, especially with regard to effective support for learners with different levels of prior knowledge. This meta-analysis summarizes the effects of scaffolding and technology use in simulation-based learning environments on facilitating a range of complex skills across domains (e.g., medical and teacher education, psychological counseling, care). In a previous meta-analysis, it has been found that the effects of instruction across domains of medical and teacher education have similar magnitude for a certain set of skills related to diagnosing; the effects increase in magnitude with the proper use of scaffolding (Chernikova et al., 2019). Other meta-analyses in the field of medical education (e.g., Cook, 2014) support the idea that simulations can

be highly effective for advancing specific motor and technical skills. However, knowledge is still scarce with regard to the effective support to advance a variety of complex skills for learners with different levels of prior knowledge. The present meta-analysis aims at advancing this research by summarizing the effects of simulation-based learning on complex skills—going beyond understanding the subject matter and performing technical tasks. In addition, this meta-analysis aims at differentiating the effects for learners on different levels of prior knowledge.

To assess the effects of instructional support, this meta-analysis adopts a scaffolding framework suggested by Chernikova et al. (2019). The framework relies on defining scaffolding as support during working on a task connected with a temporary shift of control over the learning process from a learner to a teacher or learning environment (e.g., Tabak & Kyza, 2018). The framework suggests that learners with different levels of prior knowledge would benefit from different types of scaffolding. More specifically, learners with high levels of prior knowledge would benefit more from scaffolding that affords and requires more self-regulation (e.g., inducing reflection phases), whereas learners with a low level of prior knowledge would rather benefit from more guidance (e.g., through examples).

Complex Skills in Higher Education

In higher education, students need to be prepared for their future profession, and their professional competences should involve a range of complex skills. The importance of 21st-century skills goes beyond secondary education and is also often addressed during higher and further education (e.g., P21, 2019). Critical thinking, problem solving, communication, and collaboration seem to be the most relevant skills that students should acquire during their education in addition to domain-specific knowledge and skills to be able to make professional decisions and implement solutions.

According to Mayer (1992), *problem solving* in a broader sense is cognitive processing aimed at achieving a goal when no solution path or method is obvious. It involves critical thinking, monitoring, and experimental interactions with the environment (Raven, 2000), as well as directed application of knowledge to the case or problem. Shin et al. (2003) emphasize the differences between well- and ill-structured problems. Well-structured problems present all elements of the problem, engage the application of a limited number of rules and principles and possess correct, convergent answers. A good example of such problems would be finding the predicate in a sentence, calculating drug dosage based on a patient's weight, and so on. However, the real-world problems that professionals across domains deal with are usually ill structured. In turn, they fail to present one or more of the problem elements; might have unclear goals; possess multiple solutions, solution paths, or sometimes no solutions at all; or represent uncertainty about which concepts, rules, and principles are necessary for the solution or how they are organized (Funke, 2006; Shin et al., 2003). Problem solving in this case might involve not only *diagnosing* but also *managing critical situations*. By diagnosing, we understand collecting case-specific information to reduce uncertainty (Heitzmann et al., 2019), such as diagnosing learning difficulties, identifying the cause of a health problem, or developing a course of action. By managing critical situations we understand using the set of skills required to behave in situations of

emergency or uncertainty, such as classroom management skills or emergency help with disasters or accidents.

If the solution path is already known (e.g., a complex procedure with a set of required steps needed to collect information or make a decision) and needs to be accurately followed, we would rather address it as technical/manual *performance* (an example from medical education would be completing an examination or operation, in teacher education—carrying out lesson activities according to a plan prepared in advance).

To implement the decisions and solutions to problems as well as to collect missing information, one also needs *communication skills* (e.g., to persuade others to help or to collect missing information from other people; Raven, 2000). If multiple professionals are involved in the situation and need to collaborate to solve the problem, make the decision, or take action (e.g., an emergency team), we see it as *collaboration/teamwork* skills.

To sum up, different professional domains require specific professional knowledge as a prerequisite to enable the implementation of complex skills; however, the complex skills required appear to be similar across domains. One should be able to identify the problem, analyze the context, and apply professional and experiential knowledge to make practical decisions. Fischer et al. (2014) suggested a framework of epistemic activities that is relevant to a broad range of problem-solving and decision-making procedures across domains: identifying the problem, questioning, generating hypotheses, constructing artifacts, generalizing and evaluating evidence, drawing conclusions, and communicating processes and results. In this meta-analysis, focus on learning outcomes that involve these epistemic activities, critical thinking, and problem solving are among the most important eligibility criteria.

Simulations in Educational Contexts

Simulation-based learning offers learning with approximation of practice, allows limitations of learning in real-life situations to be overcome, and can be an effective approach to develop complex skills. Beaubien and Baker (2004) define a simulation as a tool that reproduces the real-life characteristics of an event or situation. A more specific definition suggested by Cook et al. (2013) stated that simulation is an "educational tool or device with which the learner physically interacts to mimic real life" and in which they emphasize "the necessity of interacting with authentic objects" (p. 876).

What makes simulations educational tools is the opportunity to alter and adjust some aspects of reality in a way that facilitates learning and practicing (e.g., they address less frequent events, shorten response time, provide immediate feedback to the learner, etc.). Although feedback, as providing information about discrepancy of current state (or behavior) and a desired goal stat (Hattie & Timperley, 2007), plays an important role in designing simulation, there are much more opportunities for instructional support. The present research aims at exploring opportunities to provide additional information and scaffolding to the learner in detail.

The operational definition of simulation includes interaction with a real or virtual object, device, or person and the opportunity to alter the flow of this interaction

with the decisions and actions made by learners (Heitzmann et al., 2019). Thus, all types of interaction, from role plays and standardized patients to highly immersive interactions with virtual objects, can be considered simulations.

The operational definition of simulation also implies that there is critical thinking and a kind of problem solving that is present during learning and learners take an active role in the skill development processes. Simulations have many features to address the complexity of real-world situations (Davidsson & Verhagen, 2017); for example, they can involve technology aids to better resemble reality or to provide more practice or learning opportunities. However, the idea of simulation tends to focus on the reconstruction of realistic situations and the genuine interactions that participants can participate in. According to Grossman et al. (2009), simulation can be viewed as a simplified version of practice and be used to engage novices in practices that are more or less proximal to the practices of a profession. Therefore, it seems reasonable to additionally focus on the duration of simulation and authenticity in order to determine how realistic the learning environment was and how long the learners were exposed to it.

Another aspect that needs to be considered is the *type of simulation*, which can be categorized according to what or who learners interact with (real or virtual object or person). The type of simulation is related to the concept of information base (e.g., where the information for decisions comes from) as a context variable (Chernikova et al., 2019; Heitzmann et al., 2019) but provides a further categorization into real or virtual object (e.g., document, tool, model) and real or virtual person (e.g., standardized patients).

Technology use refers to the application of digital media (hardware and software) to establish a learning environment. In class, role plays, simulated discussions, and communication with standardized patients can be seen as simulations without technology use, as no software or hardware is necessary to initiate the interaction (e.g., Davidsson & Verhagen, 2017). Screen-based simulations require computer-supported interfaces and some software, which allows the interaction (e.g., Biese et al., 2009). Another type is interaction supported by combining some hardware with software, such as in a programmed mannequin (Liaw et al., 2014). One more type that requires complex technology is virtual reality, which is likely to facilitate immersion (e.g., Ahlberg et al., 2007). Empirical research on the effects of technology use on learning (comparing the use of computers in the classroom with no technology) provides some supportive evidence of small to moderate positive effects of technology use on learning and achievement (Hattie, 2003; Tamim et al., 2011). Some evidence of no particular effects of technology use comes from medical education comparing high- and low-fidelity simulations for learning (e.g., Ahad et al., 2013). Systematic empirical evidence of the effects of virtual reality is lacking due to the fact that virtual reality is rather new technology and is not yet broadly implemented in classrooms. The aim of this meta-analysis is to summarize the effects of different technologies on acquiring complex skills.

Duration of simulation refers to the time of exposure to a learning environment. In their meta-analysis, Cook et al. (2013) provided supportive evidence of the effectiveness of distributed and repetitive practice (effects above .60) in the acquisition of complex skills in medical education. However, they only focused

on comparing simulations with a duration of more than 1 day with simulations of less than 1 day. Therefore, as a next step a meta-analysis might be aimed at capturing the effects of the duration of simulations on a more detailed scale (including simulations lasting for several minutes, hours, days, weeks, or semesters).

Simulation-based learning allows reality to be brought closer into schools and universities. Learners can take over certain roles and act in a hands-on (and heads-on) way in a simulated professional context. Research has shown that full authenticity is not always beneficial for learning (e.g., Henninger & Mandl, 2000). Researchers therefore typically emphasize the opportunity to modify reality for learning purposes with simulated environments. Therefore, it is important to consider the extent to which simulation represents the actual practice in terms of demands set on the learner, the nature of the simulated situation, and the environment and/or the participants involved (e.g., Allen et al., 1991). Sometimes this relationship is addressed as fidelity of the simulation. However, a recent review by Hamstra et al. (2014) emphasized that there are severe inconsistencies in using this term in different research areas. In line with recommendations given by this group of authors (Hamstra et al., 2014), we focus on functional correspondence between a simulated scenario or the simulator itself and the context of real situations. We address this correspondence by estimating the degree of *authenticity*, and in this way, we avoid the term *fidelity* and the uncertainty related with it.

Prior Knowledge and Professional Development

Prior knowledge is an important predictor of learning success (e.g., Ausubel, 1968); it can also define the ability of learners to learn from particular materials, and use learning strategies. There are considerations suggesting that including simulations in a later phase of a higher education program after students already know theoretical concepts is promising in order to not overwhelm learners and block too much of their cognitive capacity for solving a problem in a simulation (e.g., Kirschner et al., 2006). Other considerations suggest including simulations at a very early point because this supports the process of restructuring knowledge into higher order concepts that can be used directly to solve problems (e.g., Boshuizen & Schmidt, 2008; Schmidt & Boshuizen, 1993; Schmidt & Rikers, 2007). Thus, a theoretical knowledge base would be stored in a more effective way that is directly linked to cases of application (e.g., Kolodner, 1992). Nonetheless, it seems rather obvious that learners with less theoretical prior knowledge may need more instructional guidance than more advanced learners in order to still possess enough cognitive resources for learning (e.g., Schmidt et al., 2007).

Including opportunities to apply knowledge such as in simulations in higher education programs is crucial. It seems reasonable not to assume a general answer to the question of when a simulation should be used in a higher education program. The type of simulation and the type of instructional support in relation to the prior knowledge of the learners may be more indicative and is therefore explored in the current meta-analysis.

Added Value of Instructional Support

Exposure to ill-structured problems, especially in early stages of expertise development, should be accompanied by scaffolding to maximize learning and

avoid cognitive overload, distraction, or focusing on superficial features of a situation (see the discussions of Hmelo-Silver et al., 2007; Kirschner et al., 2006). Scaffolding enables a learner to solve problems through modifying tasks and reducing possible pathways, and through hints helping the learner to coordinate the steps in problem solving or interaction (Quintana et al., 2004), by taking over some elements of learning material (Wood et al., 1976). In meta-analyses, scaffolding has been shown to have medium effects on various learning outcomes (e.g., Gegenfurtner et al., 2014).

One of the scaffolding features frequently mentioned by the researchers is the opportunity to adjust its amount. Scaffolding can be presented during or after the learning situation, be present all the time, or be added or faded gradually (e.g., Belland et al., 2017; Van de Pol et al., 2010). The recent meta-analysis, however, showed that at least in the domains of medical and teacher education, the majority of studies implementing scaffolding to foster diagnostic competences do not employ fading or adding procedures but still report positive effects (Chernikova et al., 2019).

Instructional support can be implemented in many different ways: Learners can obtain a theoretical introduction or some information on how to deal with materials in advance (knowledge convey) or they can be scaffolded in the learning environment. Learners can be guided through procedures step by step (e.g., worked examples or modeling); provided with observation scripts, checklists, or a set of rules or ways to deal with the case in question (e.g., prompts); assigned specific roles (with a prescribed course of action or goals); and asked to reflect on their own problem solving, set goals, and assess progress (e.g., inducing reflection phases).

These types of scaffolding can be positioned on the scale from high levels of instructional guidance and little need for self-regulation to develop skills to high levels of self-regulation with little instructional guidance (Chernikova et al., 2019). In this framework, examples provide solutions or model target behavior (e.g., Renkl, 2014) and can be positioned at a high level of instructional guidance and therefore less self-regulation. Reflection phases, on the other hand, allow learners to think about goals, analyze their own performance, and plan further steps (e.g., Mann et al., 2009), but they do not provide much guidance during the problem solving. Assigning roles can be viewed as prescribing a certain way of solving the problem. Prompts are scaffolds providing hints or additional information about how to handle the task in terms of the actual process required (e.g., Quintana et al., 2004) and might contain higher or lower levels of guidance. All these scaffolding measures were found to be beneficial for developing diagnostic competences (Chernikova et al., 2019). Moreover, the study found an interaction effect between types of scaffolding and prior professional knowledge, suggesting that learners with higher prior knowledge benefit more from scaffolding with less instructional guidance, while learners with low prior knowledge benefit more from scaffolding with more instructional guidance. In this meta-analysis, we aim to replicate the findings on the interaction effect for a broader set of complex skills.

Research Questions

Research Question 1: (a) To what extent can simulation-based learning environments facilitate the development of complex cognitive skills in higher

education? (b) Are the effects of simulation-based learning and scaffolding generalizable across different complex skills?

Quite strong empirical evidence supports learning through problem solving in postsecondary education in general (Belland et al., 2017; Dochy et al., 2003). Simulation in turn can be viewed as one of the ways to apply problem solving in close-to-reality settings. There is also empirical evidence in favor of using simulation in medical (e.g., Cook, 2014; Cook et al., 2013) and nursing (Hegland et al., 2017) education to facilitate learning. In line with previous research findings, we expect moderate to large effects of simulation-based learning on the development of complex skills.

Research Question 2: How do simulation features (type of simulation, technology use, duration, and authenticity) contribute to the effectiveness of a simulation-based learning environment?

We assume that type of simulation, technology use, duration, and authenticity might contribute to the involvement of learners, which in turn might have an effect on the development of complex skills. We expect small to moderate overall effects of technology use (Tamim et al., 2011), positive effects of longer duration (e.g., Cook, 2014), and higher authenticity of simulations (e.g., Hamstra et al., 2014).

Research Question 3: To what extent does instructional support contribute to the effectiveness of simulations?

Belland et al. (2017) found moderate to large effects of scaffolding in computer-based learning environments. The meta-analysis by Chernikova et al. (2019) reported positive effects of different scaffolding types (e.g., role taking, prompts, reflection phases) on the development of diagnostic competences in medical and teacher education. In line with these findings, we expect that scaffolding would have a positive effect on the advancement of complex skills beyond the effects of simulation-based learning environment. We adopt the scaffolding types categorized in the study (Chernikova et al., 2019) and expect to find a similar pattern of the effects in simulation-based learning environments and for a broader set of complex skills. We expect significant positive added value from the scaffolding compared with simulations with no scaffolding provided.

Research Question 4: (a) To what extent does the learner's prior knowledge (i.e., familiarity with the context, level of education) contribute to the effectiveness of simulation-based learning? (b) How does prior knowledge moderate the effect of different scaffolding types on different complex skills?

We expect simulation-based learning to be effective for learners with both low and high prior knowledge when supported by additional instructional measures (e.g., scaffolding). Moreover, in line with the recent meta-analysis by Chernikova

et al. (2019), we expect to find an interaction between learners' prior knowledge and the effectiveness of different scaffolding types, with the scaffolding providing high levels of guidance to be more effective for unfamiliar contexts (a lower level of education), and that relying on high levels of self-regulation to be more effective for familiar contexts (higher levels of education). We expect that the added value of the scaffolding in a simulation-based learning environment will be even more pronounced for learners with low levels of prior knowledge.

Method

Inclusion and Exclusion Criteria

The inclusion criteria were based on the outcome measures reported, research design applied, and statistical information provided in the studies. We discuss these criteria below in more detail.

Complex Skills

The studies eligible for inclusion had to focus on the facilitation of complex skills related to critical thinking, problem solving, communicating, or epistemic activities (Fischer et al., 2014) performed individually or collaboratively. Types of outcomes were first collected as mentioned in primary studies and subsequently categorized. This meta-analysis focuses only on objective measures of learning (written or oral knowledge tests, assessment of performance based on expert rating, or any quantitative measures, including but not limited to frequency of behavior or the number of procedures performed correctly). Studies that reported only learners' attitudes, beliefs, or self-assessment of learning or competence were excluded from the analysis.

Research Design

The aim of this meta-analysis was to draw causal inferences regarding the effect of a simulation-based learning environment and instructional support on the development of complex skills, so the studies eligible for the analysis had to either have an experimental or quasi-experimental design with at least one treatment and one control condition or report both pre- and postmeasures in the case of a within-subject design. The treatment condition had to include a simulation-based learning environment with instructional support measures, and the control condition should not include active participation in a simulation-based learning environment but could include other instructional methods (i.e., traditional teaching). Studies that did not report any intervention (i.e., studies on tool or measurement validation), studies that reported the comparison of multiple experimental designs (e.g., simulation with few prompts vs. simulation with many prompts, or using best-practice examples vs. erroneous examples within a simulation), and studies that did not provide any control condition (control group or pretest measures) were excluded from the analysis. Study design was used as a control variable in the analysis.

Study Site, Language, and Publication Type

Eligible studies were not limited to any specific study site. To ensure that the concepts and definitions of the core elements coded for the meta-analysis were

comparable and relevant, only studies published in English were included in the analysis. However, the origin of studies and language of conduction were not restricted. Different sources, both published and unpublished, were considered to ensure the validity and generalizability of the results. There were no limitations regarding publication year. Publication type was used as a control variable.

Effect Sizes

Eligible studies were required to report sufficient data (e.g., sample sizes, descriptive statistics) to compute effect sizes and identify the direction of scoring. If a study reported information about the pretest effect size, it was used to adjust for pretest differences between treatment and control conditions.

Search Strategies

The search terms were simulat*, competenc*, skill*, teach*, and medic* with no restriction on where the terms occur (title, abstract, descriptor, or full text); we also included experiment* OR quantitative* OR control OR quasi OR effect OR impact in the search string to focus on experimental studies testing the effects of treatment on learning. The search results were obtained on April 13, 2018, and included all articles published before April 2018 in the PsycINFO, PsycARTICLES, ERIC, and MEDLINE databases. After deleting the duplicates, the search resulted in 3,235 articles in medical and teacher education, counseling, engineering, management, and care. During abstract and full-text screening, studies that met the exclusion criteria mentioned in the previous section were excluded; all other studies were included in the next step of the screening and the analysis (Figure 1).

Coding Procedures

The coding scheme for moderators was developed based on a conceptual model developed by the research group (Heitzmann et al., 2019). First, the studies were coded for eligibility criteria using Covidence (online version; Veritas Health Innovation, 2019); the flow chart is presented in Figure 1. Coding for eligibility was done by the first author and three student research assistants using Covidence (online version; Veritas Health Innovation, 2019). If in any doubt (not obvious exclusion), the study was included for further screening. In 95% of cases, the coders agreed about eligibility in the first round. In regular meetings, the coders discussed studies for which there was uncertainty related to eligibility until complete agreement on the inclusion or exclusion of a study was achieved.

Second, within the coder training, 50% of primary studies were double coded (with an interrater agreement above .80). All discrepancies were discussed to reach the final agreement of 100%, and after agreement was reached, all the studies (including training material) were coded by the same author and student research assistants independently. All interrater agreement values can be found in the coding manual, submitted as supplemental document. For "authenticity," all the studies were double coded and the initial interrater agreement was 78% (with a Cohen's kappa of .65, due to strong imbalance in the amount of studies in each category), the disagreement was resolved through discussion, and the studies that did not provide enough information for the unambiguous decision ($N = 37$) were excluded from this part of the analysis.

```
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ 5145 references for screening │ ───→  │ 1910 duplicates removed       │
└─────────────────────────────┘        └─────────────────────────────┘
                │
                ▼
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ 3235 titles and abstracts     │ ───→  │ 2295 studies irrelevant       │
│ screened                      │        │                               │
└─────────────────────────────┘        │ No intervention, wrong        │
                │                        │ population, or not objective  │
                ▼                        │ measures of learning          │
                                         └─────────────────────────────┘
```

5145 references for screening — 1910 duplicates removed

3235 titles and abstracts screened — 2295 studies irrelevant
No intervention, wrong population, or not objective measures of learning

940 full texts screened — 720 studies excluded
- do not fit definition of simulation
- no empirical data provided
- no control group
- not an intervention study
- not about fostering complex skills
- duplicates (reporting same intervention)

220 eligible studies meeting eligibility criteria — 92 studies not providing sufficient statistical data for estimating effect sizes after sending primary authors the request for missing data

128 studies reporting 145 independent experiments and providing 409 effect estimations
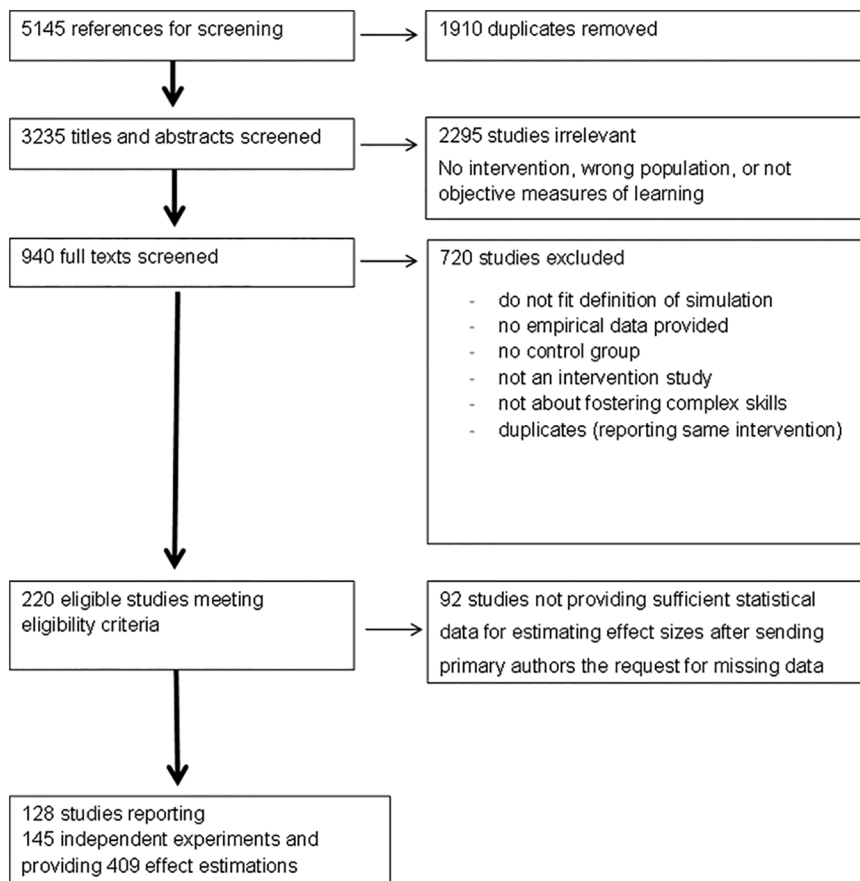
FIGURE 1. *Flow chart of study selection process.*

The *domain* was coded as medical, teacher education, or other (i.e., psychological counseling, management). *Study design* was coded as experimental (participants were randomly assigned to conditions), two-group design (no randomization took place), and one-group design (pre–post design). *Type of control group* was coded for experimental and two-group designs to distinguish between waiting condition, which did not receive any treatment, and instructed control groups, which received instructional support but no simulation. For one-group designs, the type of control was coded as "baseline," which is the level of complex skill before the intervention as measured in the pretest. Additionally, *publication year*, information about *authors* and *publication type* were retrieved from the primary studies.

Technology use was coded to address the technology support independently from the content of simulation. It was coded as "no" if no specific equipment was used to present the simulation or facilitate the interaction of learners with the

problem (e.g., role play), "computer" for screen-based simulations and interacting with virtual or real objects in a computer-supported environment, "simulator" for using specific tools (e.g., low- and high-fidelity mannequins) that learners have physically interacted with, and "virtual reality" for interacting with virtual objects or people in an immersive environment.

For *simulation type*, we focused on the source of information for learning. The simulations were categorized into *document* in the case of interaction with written information to make a decision, diagnose, and so on (e.g., disease history or students' academic achievement tests or homework), *virtual object* (interaction with a virtual object), *role play* (interaction with peers, standardized patients, or simulated students), or *live model* (interaction with real patients, students, or clients). For medical education, we also distinguished between *mannequin* (a human-like model that shows clinical symptoms) and *model* (a real object, representing the human body or its parts, that does not show clinical symptoms). In cases where multiple types of simulation were used during treatment, the code "*mixed*" was used.

Duration of simulation was collected in an open format and subsequently categorized into "very short" (lasting up to1 hour), "short" (lasting 1 or more hours, up to 1 day), "medium" (lasting 1 or several days, up to a month), or "long" (lasting more than 1 or several months).

Authenticity was coded as *low* if the simulation (task, scenario, or equipment used) resembled reality (real tasks and activities) to a low degree. Authenticity was coded as *high* if the simulation resembled reality in much detail and in different aspects (e.g., teaching to a simulated classroom, using high-fidelity simulators in medicine, whole facility simulation with different professionals involved in real-time simulation). If only one aspect of the task or situation resembled reality to a high degree and all other aspects were not presented in the way the real task requires, authenticity was coded as *selected*.

The *complex skills* were collected in an open format from the primary studies based on target learning outcomes mentioned and were subsequently categorized into the following: "diagnosing" if diagnosing was involved (e.g., diagnosing learning difficulties in teacher education or a particular disease in medical education); "technical performance" for completing complex procedures following known steps or a checklist (performing laparoscopy in medicine or implementing the teaching technique in class for teacher education); "communication skills" for assessment measures of the quality of interaction with other people; "teamwork" for outcome measures related to coordinated performance in a collaborative task; "general problem solving" for problem solving not involving diagnosing (e.g., use of argumentation, identifying or setting goals); and "management" for outcome measures related to managing critical situations (e.g., classroom management or management of critical situations in nursing or medicine).

Prior knowledge was coded in two dimensions: familiarity of context and level of education. Familiarity was coded as low if participants of the study had had little or no exposure to a similar context, and high if learners had already been exposed to a similar context. Level of education was coded as low for undergraduate and graduate students and high for postgraduate students and licensed professionals.

Scaffolding was coded as "included" or "not included" for the following categories: (1) *examples* (observing modeled behavior or example solutions), (2) *prompts* (receiving hints on how to work in simulated scenarios), and (3) *reflection phases* (thinking about the goals of the procedure, analyzing own performance and planning further steps). Additionally, knowledge convey (e.g., lecture or info session prior to simulation-based learning) was coded as "included" or "not included" to control for additional instructional support beyond the scaffolding. In the post hoc analysis, based on the initial coding mentioned above, the scaffolding was recoded to capture the combinations "no scaffolding," "examples only," "prompts only," "reflections only," "examples + prompts," "examples + reflections," "prompts + reflections," and "all included."

Statistical Analysis

Calculation of the Effect Sizes and Synthesis of the Analysis

For our analyses, we followed the procedure described by Borenstein et al. (2009) for effect size calculation, integration, and moderator analysis. First, the data on the effects of selected studies was gathered with the help of an Excel sheet. In addition to the coding of moderator variables and statistical values, the Excel sheet was used to compute Cohen's *d*, variance, and *SE* of Cohens's *d* as well as correction factor J (see Borenstein et al., 2009). Then, R Studio (Version 3.5.0., 2018) was used to calculate Hedges's *g* and perform effect aggregation and metaregression ("metafor" and "robumeta" packages). Second, as multiple studies reported multiple outcomes and used several treatment and/or control conditions, the correlated effect sizes were handled by using robust variance estimation correction coefficients, as suggested by Tanner-Smith et al. (2016). Third, the effect sizes were controlled by pretest differences as these differences may increase the share of random effects variance. As prior knowledge has to be regarded as an important predictor of knowledge at posttest (e.g., Ausubel, 1968), it has to be controlled for if possible. If available, pretest data were used to adjust the effect sizes by subtracting the pretest effect sizes from the posttest effect sizes and adding up the variances of both effects. Fourth, preliminary analyses were performed to identify systematic bias among effects from primary studies. Fifth, a random effects model was used to address the research questions. Confidence intervals were employed to assess the significance of an effect. Heterogeneity estimates (Q-statistics) were utilized to determine the variance of the true effect sizes between studies (tau) and the proportion of this variance that could be explained by random factors (I^2). The thresholds suggested by Higgins et al. (2003) were used to interpret the I^2 (25% for low heterogeneity, 50% for medium, 75% for high heterogeneity).

Assessment of Publication Bias

The recent simulation study by Carter et al. (2017) compared the effectiveness of different techniques to estimate and correct for publication bias under different conditions (i.e., high heterogeneity). They suggest not relying only on the results of the random effects model if the probability of publication bias is high. We expect high effects of simulation-based learning (i.e., Cook, 2014) and do not expect high publication bias; however, we expect high heterogeneity of the effects,

which according to Carter et al. (2017) can hinder some types of analysis (e.g., p-curve analysis). We have a relatively small sample of empirical studies and apply a range of techniques to identify possible biases (Egger's test, trim and fill) and improve the generalizability of the results (Sterne & Egger, 2001). If no publication bias were detected, we would rely on random effects model estimation for the effect sizes with robust variance estimation correction.

Results

Results of Literature Search

The 145 eligible studies (from 128 articles published in the period 1979–2018) provided 409 effect estimations (Figure 1). The total sample consisted of 10,532 participants. Most of the studies come from medical education (126 studies), while studies in teacher education are represented by seven independent studies and other domains by 12 independent studies. Most studies focused on general problem solving (51) or the technical performance of a particular complex procedure (56); the other outcomes were communication skills (24), diagnosing (18), managing critical situations (18), and collaboration and teamwork (5). Some studies reported more than one complex skill as learning outcomes.

Out of 409 effects of simulation-based learning, only 270 included complete information with no missing codes on instructional support measures used within the simulation. In 12% of treatments, the simulation was not accompanied by any additional instructional support, while 25% of simulations were accompanied by knowledge convey (i.e., lectures or other expository forms of instruction). It is worth noting that a small number of simulations reported using a single scaffolding type only: 6% used examples with no other support measures, 3% used simulations with additionally induced reflection phases, and less than 1% used solely prompts to support simulation. The most frequent combinations of instructional support measures were knowledge convey together with examples (82 effects), knowledge convey with reflection phases (62 effects), and examples with reflection phases (43 effects). However, the analysis identified that there is a considerable amount of missing data indicating that the instructional support measures were not mentioned explicitly or in sufficient detail in the description of the treatment in primary studies. Additionally, almost all of the studies reported some kind of feedback that participants received from the learning environment or the instructor during or after the simulation, which was not explicitly coded and therefore stayed beyond the focus of the current analysis.

Quality Assessment and Preliminary Analysis

The procedures targeted at assessing the quality of data coming from primary studies (e.g., no linear relationship between effect size and standard error, symmetry of funnel plot) and the generalizability of the summary and moderator effects found in the meta-analysis indicated no evidence of publication bias or questionable research practices (see Figure 2 and Table 1). The metaregression on control variables (year of publication, publication type, study design, type of control, domain) showed that these factors do not explain any statistically significant amount of variance between study effects (*p* values above .05).
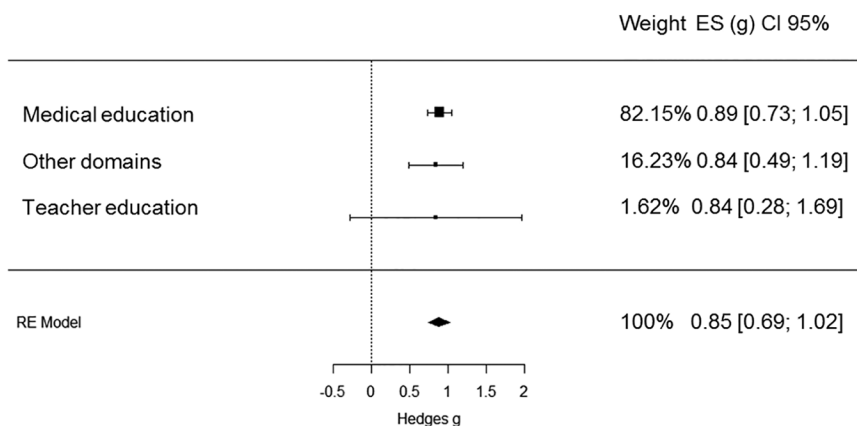
	Weight	ES (g)	CI 95%
Medical education	82.15%	0.89	[0.73; 1.05]
Other domains	16.23%	0.84	[0.49; 1.19]
Teacher education	1.62%	0.84	[0.28; 1.69]
RE Model	100%	0.85	[0.69; 1.02]

-0.5   0   0.5   1   1.5   2

Hedges g

FIGURE 2. *Forest plot of the overall effect of simulations on the acquisition of complex skills across domains.*

Overall Effect of Simulation-Based Learning on Complex Skills (Research Question 1)

With regard to Research Question 1, simulation-based learning had a large positive effect on fostering complex skills compared with conditions (1) without intervention (waiting control: $g = 1.02$, $SE = 0.30$, $N = 16$); (2) with differently instructed control: $g = 0.82$, $SE = 0.13$, $N = 53$); or compared with (3) baseline ($g = 0.88$, $SE = 0.10$, $N = 76$). As there were no statiscally significant differences between three control conditions, the overall effect was estimated: $g = 0.85$, $SE = 0.08$, $N = 145$. As expected, the analysis also identified high heterogeneity between studies: $Q (409) = 4213.93$, $p < .0001$; $\tau^2 = 1.2$; $I^2 = 95.86\%$. This heterogeneity could not be explained by control variables (year of publication, publication type, study design, type of control, domain). The effect sizes found in individual studies, weights, and confidence intervals, as well as the summary effect from the random effects model estimation, are presented in Figure 2 and organized by domains. A funnel plot of effect size distribution and standard errors is presented in Figure 3.

Communication and collaboration skills (teamwork) are only moderately facilitated by simulation-based learning ($g = 0.44$, $SE = 0.15$, and $g = 0.50$, $SE = 0.08$, respectively), followed by situation management ($g = 0.72$, $SE = 0.30$), diagnostic skills ($g = 0.82$, $SE = 0.21$), and problem solving ($g = 0.88$, $SE = 0.11$); the highest effects of simulations reported are on technical performance ($g = 1.06$, $SE = 0.15$). The number of studies in each group can be found in Table 1.

Features of Simulation (Research Question 2)

Simulation Type

Simulation-based learning had greater effects when presented in the form of live simulations with real patients ($g = 2.27$, $SE = 1.04$, $N = 3$, medical education only), followed by hybrid simulations, where several simulation types were

TABLE 1

Effects of simulation features, prior knowledge, and instructional support on acquisition of complex skills.

Summary effect (random effects model)	p of Q	Q (p)	g	95% CI	N (k)	τ^2	I^2	EV (N)
Simulation vs. control	<.001	4213.93	0.85	[0.69, 1.02]	10,532 (145)	1.21	95.86%	1 (98)

Moderator variables	Significance of moderator		Effect size			Heterogeneity		Quality assessment
	p of Q	Q between	g	95% CI	n(k)	τ^2	I^2	EV(N)
Complex skills	<.001	167.61						
Communication skills			0.44	[0.17, 0.72]	2,493 (27)	0.46	91.03%	1 (17)
Diagnostic skills			0.82	[0.41, 1.22]	911 (18)	0.76	92.02%	1 (14)
General Problem solving			0.88	[0.68, 1.08]	6,010 (58)	0.47	91.23%	1 (39)
Management of situation			0.72	[0.14, 1.31]	2,543 (21)	1.04	97.15%	1 (14)
Teamwork/skills			0.50	[0.32, 0.68]	810 (5)	0.03	66.07%	1 (3)
Technical performance			1.06	[0.75, 1.37]	2,933 (63)	1.25	91.10%	1 (43)
Simulation features								
Type of simulation	<.001	852.62						
Document			0.31	[0.07, 0.56]	847 (15)	0.30	82.60%	1 (7)
Virtual object			0.75	[0.47, 1.03]	3,199 (45)	0.80	95.49%	1 (32)
Role play (SP)			0.63	[0.38, 0.89]	1,934 (26)	0.35	87.02%	1 (20)
Mixed (more than one type)			1.56	[0.90, 2.22]	936 (13)	1.18	95.98%	1 (13)
Live model (medical only)			2.27	[1.67, 2.86]	108 (3)	0.08	18.56%	1 (3)
Mannequin (medical only)			0.96	[0.60, 1.31]	3,015 (30)	1.10	94.41%	1 (22)
Model (medical only)			0.79	[0.33, 1.26]	589 (18)	0.81	87.56%	1 (9)

TABLE 1 (CONTINUED)

Moderator variables	Significance of moderator		Effect size		n(k)	Heterogeneity		Quality assessment
	p of Q	Q between	g	95% CI		τ²	I²	EV(N)
No			0.74	[0.53, 0.96]	4,895 (54)	0.58	91.16%	1 (39)
Computer-supported			0.68	[0.40, 0.97]	2,578 (43)	0.68	91.67%	1 (24)
Simulator (medical only)			1.07	[0.66, 1.47]	1,312 (26)	1.26	94.19%	1 (21)
Virtual reality			0.85	[0.31, 1.39]	1,377 (20)	1.25	97.67%	1 (17)
Duration of simulation	<.001	439.42						
Very short (up to 1 hour)			0.65	[0.19, 1.10]	2,210 (26)	1.27	97.22%	1 (17)
Short (up to 1 day)			0.81	[0.65, 0.97]	5,496 (87)	0.62	89.16%	1 (61)
Medium (up to 1 month)			0.80	[0.54, 1.07]	459 (11)	0.13	59.62%	1 (9)
Long (more than 1 month)			1.31	[0.00, 2.64]	519 (4)	1.37	91.13%	1 (3)
Authenticity	<.001	265.58						
Low			0.58	[0.28, 0.88]	1,598 (26)	0.58	89.57%	1 (15)
Selected			0.69	[0.00, 1.41]	268 (6)	1.04	89.12%	1 (5)
High			0.86	[0.65, 1.07]	6,164 (76)	0.83	95.21	1 (56)
Instructional support								
Knowledge convey	<.001	32.12						
Present			0.87	[0.69, 1.05]	7,925 (98)	0.71	93.95%	1 (75)
Not present			0.72	[0.33, 1.10]	2,323 (46)	1.29	93.15%	1 (30)
Scaffolding	<.001	237.91						
No scaffolding			0.88	[0.64, 1.12]	4,824 (58)	0.68	92.66%	1 (42)
Examples only			0.66	[0.22, 1.10]	1,046 (27)	1.67	93.04%	1 (17)
Prompts only			0.44 ns	[−0.18, 1.07]	554 (11)	1.18	90.94%	1 (4)

(continued)

515

TABLE 1 (CONTINUED)

Moderator variables	Significance of moderator		Effect size		n(k)	Heterogeneity		Quality assessment
	p of Q	Q between	g	95% CI		τ²	I²	EV(N)
Examples + Prompts			1.60	[0.87, 2.34]	187 (4)	0.11	30.85%	1 (4)
Examples + Reflection			0.95	[0.36, 1.54]	1,677 (15)	1.11	97.98%	1 (12)
Prompts + Reflection			0.10 ns	[−0.27, 0.48]	634 (8)	0.21	73.53%	0 (3)
All combined			1.34 ns	[−0.33, 3.02]	90 (2)	1.29	87.37%	Not applicable
Examples present			0.88	[0.54, 1.21]	3,202 (44)	1.33	96.39%	1 (31)
Examples not present			0.81	[0.65, 0.97]	8,312 (101)	0.61	91.88%	1 (87)
Prompts present			0.65	[0.20, 1.10]	1,604 (25)	0.98	90.04%	1 (13)
Prompts not present			0.92	[0.73, 1.10]	8,747 (121)	0.87	94.71%	1 (95)
Reflection phases present			0.78	[0.46, 1.10]	3,210 (39)	0.74	95.46%	1 (29)
No reflection phases			0.80	[0.57, 1.04]	5,298 (80)	1.12	93.83	1 (54)
Prior knowledge								
Familiarity of context	<.001	614.32						
Unfamiliar (low prior knowledge)			0.67	[0.40, 0.94]	5,938 (61)	1.08	96.09%	1 (44)
Familiar (high prior knowledge)			0.83	[0.65, 1.02]	2,511 (63)	0.50	86.05%	1 (48)
Mixed group			1.21	[0.50, 1.93]	1,227 (15)	1.16	97.34%	1 (12)
Level of education	<.001	329.58						
Low (undergraduate and graduate)			0.74	[0.54, 0.94]	7,143 (74)	0.71	94.66%	1 (56)
High (postgraduate and in-service)			0.91	[0.67, 1.16]	3,400 (72)	0.94	93.16%	1 (55)

(continued)

TABLE 1 (CONTINUED)

| Moderator variables | Significance of moderator | | Effect size | | | Heterogeneity | | Quality assessment |
	p of Q	Q between	g	95% CI	n(k)	τ²	I²	EV(N)
Familiar context								
Examples			0.85	[0.55, 1.15]	880 (21)	0.63	86.51%	1 (17)
Prompts			0.33 ns	[−0.40, 1.07]	330 (7)	0.37	76.49%	1 (3)
Reflection phases			0.74	[0.48, 1.00]	778 (19)	0.19	71.41%	1 (14)
Unfamiliar context								
Examples			0.72	[0.16, 1.27]	1,993 (19)	1.45	97.82%	1 (17)
Prompts			0.85	[0.19, 1.50]	1,091 (15)	1.39	92.59%	1 (9)
Reflection phases			0.49	[0.17, 0.81]	2,017 (18)	0.37	94.11%	1 (11)
Low level of education								
Examples			0.88	[0.42, 1.35]	1,689 (18)	0.93	97.28%	1 (16)
Prompts			0.74	[0.08, 1.39]	1,011 (17)	1.08	92.11%	1 (8)
Reflection phases			0.52	[0.23, 0.80]	2271 (21)	0.37	93.79%	1 (15)
High level of education								
Examples			0.85	[0.38, 1.32]	1,120 (26)	1.59	93.22%	1 (18)
Prompts			0.50 ns	[−0.08, 1.08]	346 (9)	0.83	84.19%	1 (3)
Reflection phases			1.10	[0.52, 1.68]	763 (18)	0.92	91.42%	1 (13)

Note. CI = confidence interval; EV = evidential value; SP = standardized patient; ns = not significant.

517
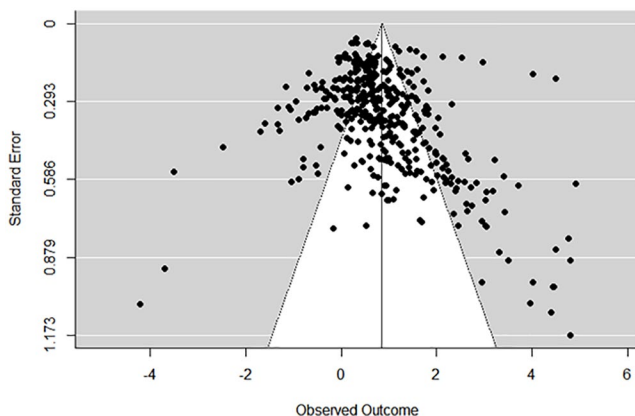
FIGURE 3. *Funnel plot of the overall effect of simulations on the acquisition of complex skills.*

used during learning phases ($g = 1.56$, $SE = 0.17$, $N = 13$). In medical education, mannequins were also highly effective ($g = 0.96$, $SE = 0.11$, $N = 30$). Role plays and using virtual objects had moderate effects on learning ($g = 0.63$, $SE = 0.12$, $N = 26$, and $g = 0.75$, $SE = 0.09$, $N = 45$), and simulations based on documents, although often highly resembling actual tasks (X-rays of patients, students' homework) were the least effective ($g = 0.31$, $SE = 0.15$, $N = 15$).

Technology Use

Higher levels of technology support during simulation were associated with greater effects on learning outcomes: simulators (e.g., programmed mannequins) in medical education ($g = 1.07$, $SE = 0.18$, $N = 26$) and virtual reality across domains ($g = 0.85$, $SE = 0.24$, $N = 20$) were more effective than screen-based simulations ($g = 0.68$, $SE = 0.13$, $N = 43$) and simulations that did not implement technology support ($g = 0.74$, $SE = 0.11$, $N = 54$).

Authenticity

Simulations that resembled reality at the low level had an effect of $g = 0.58$, $SE = 0.18$, $N = 26$, while high authenticity simulations, which represented all aspects of highly realistic situations, had an effect of $g = 0.86$, $SE = 0.10$, $N = 76$. Simulations that represented one aspect of a situation in a highly realistic way, but all other aspects were less realistic (selected authenticity), also had positive effects $g = 0.69$, $SE = 0.40$. However, the data for this type of authenticity came from a relatively small sample of studies in medical education ($N = 6$) and was highly heterogeneous.

As post hoc analysis, we looked at the interaction between authenticity and familiarity of context (learners' prior knowledge). For unfamiliar contexts, high authenticity simulations had more value ($g = 0.74$, $SE = 0.16$) than low authenticity simulations ($g = 0.57$, $SE = 0.28$). For familiar contexts, high authenticity

simulations also had more value (g = 0.92, SE = 0.14) than low authenticity simulations (g = 0.57, SE = 0.19). No interaction was found between authenticity and level of education as another indicator of prior knowledge. There was an insufficient number of studies to estimate the effects of authenticity for mixed groups.

Duration of Simulation

Very short simulations lasting less than an hour had an effect of g = 0.65, SE = 0.20. Longer simulations were associated with higher effects: simulations lasting for several hours (up to a day)—g = 0.81, SE = 0.09; simulations lasting for several days (up to a month)—g = 0.80, SE = 0.18. A few simulations lasted for more than a month and had an effect of g = 1.31, SE = 0.31, but the number of studies (N = 4) reporting simulations with an extended duration was not sufficient for conclusive results.

Added Value of the Scaffolding (Research Question 3)

To evaluate the added value of scaffolding, the effects of simulations with and without particular scaffolding types were compared. When examples were present (g = 0.88, SE = 0.17), the effects of simulations were descriptively higher than without examples (g = 0.81, SE = 0.09); however, the difference was not statistically significant. The examples across all domains were usually represented by live or recorded demonstrations of how to deal with the simulated environment, particular tool, or situation (e.g., Chen et al., 2015; Damle et al., 2015; Overbaugh, 1995). Some studies mentioned using correct or positive examples together with erroneous or negative examples (e.g., Douglas et al., 2016), but more commonly, only demonstrations of correct target behaviors were used.

The effects of simulation when prompts were presented (g = 0.65, SE = 0.23) were significantly lower than in the absence of prompts (g = 0.92, SE = 0.09). The prompts in the primary studies were represented by short textual hints within the simulation environment, which suggested actions or allowed to revisit the conceptual level during the exploration of the simulation (e.g., Dankbaar et al., 2016; Kumar & Sherwood, 2007); another form of prompting are questions that may lead the learner to further actions (Alfred & Chung, 2011).

The presence of reflection phases had no added value, while the effects with reflection phases (g = 0.78, SE = 0.16) and without reflection phases (g = 0.80, SE = 0.12) were similar. The reflection phases in the medical context introduced in the primary studies usually involved reflecting on positive and negative aspects or strategies used in the demonstration, in the own performance or a peer's performance (e.g., Alinier et al., 2006; Cuisinier et al., 2015; Douglas et al., 2016). These phases were held during briefing and debriefing sessions and usually implied that insights from these reflections can be used for further simulation trials or at least to improve one's performance in a final assessment. In nonmedical contexts, reflection phases were focused on asking learners (1) to write a report, documenting the process and problems occurring and evaluating own actions (e.g., Newell & Newell, 2018), or (2) to fill in worksheets, reflecting on their actions in the simulated environment and the consequences of these actions (e.g.,

Girod & Girod, 2006) or to reflect on the case scenario and discuss the action plan with peers and supervisors (e.g., Broadbent & Neehan, 1971).

To summarize, the comparison of presence versus absence of particular scaffolding types did not support our hypothesis about an added value of scaffolding; the effects were relatively high in the cases of presence and of absence of examples, prompts, and reflections. Post hoc analysis was performed to clarify the high heterogeneity in the effects, which could have led to a lack of significant differences in "included" versus "not included" scaffolding types.

Post hoc analysis also found that treatments with no scaffolding explicitly mentioned in the description were also connected with high effects of learning ($g = 0.88$, $SE = 0.11$), partly due to knowledge convey. But treatments with neither scaffolding nor knowledge convey ($N = 19$) also resulted in relatively high learning outcomes ($g = 0.68$, $SE = 0.21$). Furthermore, different types of scaffolding were usually combined within the study, and some combinations were more effective than others, partly supporting the hypothesis about the added value of scaffolding. For example, examples were often ($N = 15$) combined with reflection phases with the effect of ($g = 0.95$, $SE = 0.17$). Combinations of examples and prompts showed very high effects in a few studies ($N = 4$) in medical education ($g = 1.60$, $SE = 0.37$). If only prompts were used as scaffolding ($g = 0.44$, $SE = 0.32$) or combined with reflection phases ($g = 0.10$, $SE = 0.19$), no significant learning effects were found. Combinations of all scaffolding types ($N = 2$) resulted in very heterogeneous results that did not reach statistical significance ($g = 1.34$, $SE = 0.86$). To sum up, the hypothesis of the added value of scaffolding can be partly supported by post hoc analysis (see Table 1).

Prior Professional Knowledge

With regard to learners' prior knowledge, subgroup analyses (Table 1) indicated that learners showed a lower increase of complex skills in an unfamiliar context ($g = 0.67$, $SE = 0.10$) than in a familiar context ($g = 0.83$, $SE = 0.13$), but overall learners in both contexts benefited from simulation-based learning. When learners with lower and higher prior knowledge were combined in the same group (mixed group), even higher learning outcomes were reached ($g = 1.21$, $SE = 0.36$). If the level of education was taken as a measure of prior knowledge, learners both on a low level of education ($g = 0.74$, $SE = 0.11$) and on a high level of education ($g = 0.91$, $SE = 0.07$) improved their skills through simulation-based learning with a similar pattern (i.e., learners with higher prior knowledge benefited more from simulations).

Interaction Between Scaffolding Types and Prior Knowledge and Experience

There was a significant interaction effect found between prior knowledge and the effectiveness of different scaffolding types.

Familiarity of Context as an Indicator of Prior Knowledge and Experience

In a familiar context, examples have no added value, but neither do they hinder learning. Learners with higher prior knowledge (as defined by the familiarity with the context) learn equally well if examples are presented ($g = 0.85$, $SE = 0.12$) or not ($g = 0.83$, $SE = 0.15$). In unfamiliar context (learners have little prior

knowledge of what is learned), examples have more added value; however, the difference between effects if examples are presented ($g = 0.72$, $SE = 0.28$) or not presented ($g = 0.65$, $SE = 0.14$) does not reach statistical significance.

In contrast, introducing prompts in a familiar context was not beneficial for learning ($g = 0.33$, $SE = 0.38$). If no prompts were presented, the average effect of simulation in a familiar context was $g = 0.96$, $SE = 0.10$. Prompts had a significant positive effect on learning in an unfamiliar context ($g = 0.85$, $SE = 0.33$) compared with no prompts ($g = 0.63$, $SE = 0.16$).

Reflection phases induced by educators were more beneficial in familiar ($g = 0.74$, $SE = 0.15$) than in unfamiliar contexts ($g = 0.49$, $SE = 0.21$). The difference failed to reach statistical significance ($p = .13$), though.

The mixed group had an insufficient number of studies to perform the analysis of interaction with scaffolding types.

Level of Education as an Indicator of Prior Knowledge and Experience

For postgraduate learners, the presence of examples ($g = 0.85$, $SE = 0.20$) showed a smaller effect than if no examples were provided ($g = 1.00$, $SE = 0.11$). Thus, postgraduate learners had a pattern different from undergraduate and graduate learners, who benefited more from the presence of examples ($g = 0.88$, $SE = 0.27$) than if no examples were provided ($g = 0.80$, $SE = 0.11$). The differences, however, did not reach statistical significance due to high heterogeneity within the conditions where examples were present versus where examples were absent.

Similarly, introducing prompts for a high (postgraduate learners and practitioners) level ($g = 0.50$, $SE = 0.36$) was not beneficial for learning compared with simulations without prompts ($g = 0.91$, $SE = 0.11$). For low-level (undergraduate and graduate) learners, there was no statistically significant difference between the prompts ($g = 0.74$, $SE = 0.24$) and no prompts ($g = 0.76$, $SE = 0.09$) condition; however, introducing prompts was related to higher learning outcomes in the high-level group.

Reflection phases were highly beneficial for postgraduate learners ($g = 1.10$, $SE = 0.16$) compared with no reflection phases ($g = 0.86$, $SE = 0.14$). In contrast, graduate and undergraduate learners had better learning outcomes when no reflection phases were used ($g = 0.81$, $SE = 0.14$) than in the presence of reflection phases ($g = 0.52$, $SE = 0.13$).

In post hoc analysis, we analyzed only the treatments with (1) examples as the only scaffolding method ($N = 27$), (2) prompts as the only scaffolding method ($N = 11$), (3) reflections as the only scaffolding method ($N = 15$), and (4) a combination of examples and reflections as the most frequent combination ($N = 15$). Other combinations did not include a sufficient number of studies for the analysis.

For low-education-level learners, examples were more beneficial ($g = 1.15$, $SE = 0.58$, $N = 9$) than for learners with a high level of education ($g = 0.56$, $SE = 0.25$, $N = 18$).

For low-education-level learners, prompts were more beneficial ($g = 0.69$, $SE = 0.61$, $N = 7$) than for learners with a high level of education ($g = 0.14$, $SE = 0.08$, $N = 5$), but no significant effects for learning were found in either group.

For low-education-level learners, reflections were more beneficial ($g = 1.13$, $SE = 0.23$, $N = 7$) than for learners with a high level of education ($g = 0.69$, $SE = 0.15$, $N = 8$). Moreover, it was the most beneficial scaffolding for learners with a high level of education compared with the other two (examples and prompts).

The example–reflection phase combination was highly beneficial for learners with a high level of education ($g = 1.71$, $SE = 0.59$, $N = 8$), but it also had a positive effect on learners with a low level of education ($g = 0.48$, $SE = 0.19$, $N = 7$).

Discussion

The results of this meta-analysis show that simulation-based learning has large positive overall effects on the advancement of a broad range of complex skills and across a broad range of different domains in higher education. The size of the effect of simulations on learning even exceeds the expectedly large influence of the learners' prior knowledge. The effect size is still very large when simulation-based learning is compared with different kinds of instruction instead of "real" control groups, including waiting controls. There is only a very small number of instructional methods for which these relations hold true. These include feedback and formative assessment (see Hattie, 2003; Hattie & Timperley, 2007), which already point to possible interpretations of effects that large. One of the issues in higher education is the lack of feedback in the context of complex authentic activities. Simulations typically address exactly this issue. They often entail providing information to the learner on the discrepancy of currently observable competence indicators and a desired competence goal, which is one of the most common definitions of feedback in the context of learning (Hattie & Timperley, 2007). The potential of simulations for learning has been known for a while in medical education (e.g., Cook, 2014) but is now increasingly transferred (and sometimes reinvented) to other domains of higher education (see Heitzmann et al., 2019). This meta-analysis provides supportive evidence that the large effects of simulation-based learning found for medical knowledge and skills do generalize across domains.

But simulations are more than just feedback as they provide opportunities for meaningful applications of knowledge to professional problems (Grossman et al., 2009). Simulated problems may be tailored to the needs of learners as an approximation of practice and are thus probably often more effective than real practice.

With regard to types of simulation, the analysis shows that combining several types of simulation over the treatment time—for example, role play with practice on a model (Dumont et al., 2016) or virtual reality (Lehmann et al., 2013)—might have greater effects on learning. We have also identified that some types of simulation are more frequently used to target particular complex skills. For example, communication skills are frequently facilitated through role plays, whereas technical performance is frequently addressed by using a simulator or virtual reality. Therefore, we would like to emphasize that the simulation type should not be viewed independently of target skills and instructional support quality. The type of simulation depends a lot on the learning context (e.g., radiologists have to work with images, teachers with students' tests), but providing different types of simulations can be beneficial across domains.

The present meta-analysis included different types of complex skills as outcome measures. The analysis yielded evidence of differences with respect to facilitating effects that are considerably bigger than the differences between the effects of the domains involved. The biggest effect sizes were obtained for tasks related to technical performance that mainly come from medicine (Araújo et al., 2014; Banks et al., 2007), followed by problem-solving and -diagnosing skills. These findings emphasize that if the simulation requires the coordinated use of different mental modes and abilities—for example, motor and sensory skills together with reasoning—the learning gains are larger than for simulations that require the involvement of fewer skills. Despite these differences, simulations had effects that can be categorized as large positive effects for all but one type of skill. The exception is teamwork, where the meta-analysis found a medium positive effect only. This in turn may be due to the high complexity in the case of real team training. Another explanation of low effects is that it might be difficult to find ways to further improve social skills, as they are by far the most trained skills we possess.

There has been a long debate on political and societal levels around the question of an added value of technologies for learning (cf. Rogers, 2001). According to this meta-analysis, simulations still have substantial effects if no digital technology is used at all. Typical computer-and-screen-based simulations do not outperform well-organized no-tech role plays and simulated patients or simulated students. Both of them, the technology-enhanced and the no-tech variants do have large effects. However, some more recent technologies that enhance sensory perception (e.g., virtual reality, full-scale simulators) seem to make a difference. With more studies and better theory, it will be possible to identify features and dimensions of these technologies that are responsible for greater learning gains.

Another main finding of this meta-analysis is certainly that simulations with an overall high authenticity do have greater effects than simulations with a lower authenticity. However, it is also very interesting that even simulations with low authenticity still have large effect sizes, exceeding those of many other forms of instruction. This is encouraging for higher education practice as high-authenticity simulations are sometimes very expensive and time-consuming to build. At least for learners with some experience of the real situation, a reduced version might do just as well in low- as in high-authenticity simulations. Moreover, simulations aimed at high authenticity for only one or a small number of objects and processes are associated with effects similar to the effects of simulations aimed at high authenticity with respect to all situational parameters. This can be taken as supportive evidence of the approximation-of-practice approach (Grossman et al., 2009), claiming that the real advantage in simulations is the reduction of task complexity to levels a learner can handle.

A more practical question regarding the use of simulations in higher education concerns the extent to which their effects depend on prior knowledge. In other words, are simulations better suited for beginning students or for more advanced students? In this meta-analysis, the overall effects are large for both familiar and unfamiliar contexts as well as lower and higher levels of university education. However, the effects are greater if learners are unfamiliar with the context and the task. Taken together, these findings seem to indicate that even more advanced studies in higher education enable effective simulation of professional situations

of which learners do not have prior experience. This pattern of finding does not support the claim that simulations as forms of problem-based learning are only applicable in later phases in higher education when learners are familiar with the relevant concepts and procedures (see Dochy et al., 2003).

Based on the findings by Belland et al. (2017) and Chernikova et al. (2019), we expected significant positive effects of the scaffolding. However, a surprisingly small additional effect to the large effects of simulations can be attributed to scaffolding. One explanation could lay in the nature of simulation-based environments, which might already include some levels of instructional support, which is built-in in the scenario (e.g., feedback), this would also explain lower effects than the ones found by Belland et al. (2017) when comparing scaffolding with no scaffolding conditions. For the instructional support, we did not find a single pattern for effective simulation. The very same pedagogies were a success for some complex skills, while simultaneously being a failure for the other skills. We have also found that simulation-based learning implemented in primary studies has strong effects, but we admit that presenting an opportunity to interact with learning material will not improve learning by default. Additional instruction does not seem to add much beyond the effects of simulation; however, in some cases it does. A meta-analysis is not the right method to deliver detailed explanations for these exceptions. Here we need more primary studies. One contributing factor may be that in many simulations learners can find out the correct strategy themselves, by trial and error if needed. A more fine-grained analysis in the primary studies would be needed to test hypotheses stating that learners prefer trying without help instead of using assistance (see Aleven et al., 2003) also for the context of simulations in higher education.

Another important question of this meta-analysis targeted the additional instructional support and the extent to which the effects of this support depended on individual learning prerequisites, in particular, learners' prior knowledge.

The findings suggest that if learning prerequisites are not considered at all, one could even conclude that scaffolding does not make a real difference. Moreover, scaffolding is even associated with dysfunctional learning processes in some studies. The picture changes once we take the moderating effects of prior knowledge into account. This may be seen as trivial, as the training wheel effects of scaffolding had been established a long time ago (Carroll & Carrithers, 1984). The training wheels keep learners away from possible errors and their consequences, which is definitely beneficial at early stages of learning. However, the findings of this meta-analysis extend this established perspective on scaffolding. The findings support the claim made elsewhere (Chernikova et al., 2019) that different types of scaffolding rather than their presence or nonpresence have a kind of effectiveness curve in relation to learners' different levels of prior knowledge. Examples and prompts have better effects for learners with low prior knowledge, whereas reflection phases have their highest effectiveness with high prior knowledge.

Put more generally, a certain type of scaffolding may work optimally in interaction with a specific level of prior knowledge, high or low. However, whereas some types of scaffolding just lose their effectiveness with respect to the other knowledge level, others even have detrimental effects for the "nonfitting" prior knowledge level. The latter effects are known as "expertise reversal effects" from

cognitive load research (Kalyuga et al., 2003). In this meta-analysis, we found indications of reversal effects for prompts and for reflection phases. Prompts had their optimal effectiveness in unfamiliar contexts and detrimental effects in familiar contexts. For reflection phases, optimal effectiveness was given for postgraduate students and in familiar contexts, whereas graduate and undergraduate students learned better if no reflection phases were implemented.

One possible explanation for why this meta-analysis did not find a negative effect for examples might be that many of the studies included in the sample offered examples as additional options to problem solving. They did not replace problem solving with examples. Thus, more advanced learners may simply not have chosen to use them. In contrast, reflection phases were typically implemented in an intrusive way and made the learners interrupt their problem solving for some time. Prompts appeared during problem solving, attracting the learners' attention at least for some of the time needed to read and decide that the prompt was irrelevant. Of course, this suggested model of optimal effectiveness of different scaffolding types depending on learners' prior knowledge and experience needs to be put to the test in primary studies.

Limitations

Simulations are broadly used in different domains to facilitate different kinds of content knowledge and skills. The current meta-analysis puts a particular focus on learning of complex skills connected with interaction with other people, seen as complex systems (on physiological, cognitive, psychological, social, or ethical levels), and its findings do not straightforwardly generalize to other domains like science, mathematics, engineering, or informatics. One of the concerns for generalization is that a large body of the STEM research has been conducted in secondary rather than higher education, and the simulations are often used (1) to advance knowledge and skills related to interaction with mechanisms or abstract systems or (2) to understand complex concepts and their interrelation.

A large proportion of the findings in this meta-analysis comes from medical education. Although the findings and the magnitude of the effects are similar in different domains and we additionally performed sensitivity analysis to ensure the generalizability of findings across domains, some caution should be taken in interpreting results for other domains, especially with regard to the effects of technology, the type of simulation, outcome measures, and some other moderators. We hope that this analysis will instigate future studies in other domains, such as teacher education to better understand and to realize the enormous potential as well as the potential pitfalls that come with simulation-based learning.

Furthermore, there were some limitations caused by the characteristics of some of the primary studies included. First, a large part of the studies provided relatively little description of the treatment, which was insufficient for coding of some moderators, as well as differentiating on a finer level between types of reflections (reflecting on the simulated scenario vs. reflecting on own reasoning, modeling vs. worked examples or different types of prompts, e.g., cognitive, meta-cognitive). Second, many studies implemented multiple instructional support measures during one treatment making it difficult to determine the effects of a specific measure.

The study had a particular focus on the effects of different scaffolding types within a self-regulation framework (see Chernikova et al., 2019), which might have resulted in leaving out of scope some other, potentially relevant instructional support, measures like providing feedback or changing the amount of instruction during the treatment (e.g., fading or adding instruction).

The effects of combinations of instructional support could only partially be investigated due to the lack of primary studies, or missing data about scaffolding use in the studies included in the analysis. So while our study demonstrated that simulation-based learning has large positive overall effects on the advancement of a broad range of complex skills, as compared with no simulation, a necessary next step will be to directly compare different types of simulations and scaffolds with each other. However, the current body of existing research may yet not suffice for a meta-analytic approach to these comparisons.

Another limitation is the way prior knowledge was assessed in the meta-analysis. The approach of estimating learners' prior knowledge through familiarity of context and level of education proved interesting and fruitful results, but it also has some drawbacks. The level of education was more frequently indicated in the primary studies' descriptions. However, there were familiar and unfamiliar topics presented on all levels. Thus, level of education represented overall experience with learning rather than actual prior knowledge. Familiarity of context, in contrast, addresses prior knowledge of subject matter but is rarely described in the primary studies. Familiarity of context also does not directly address expertise and experience. Thus, there is still room for better operationalization of prior knowledge to explain parts of the remaining high heterogeneity: We were only able to explain 4% of the heterogeneity with the one we used.

One more limitation of the current meta-analysis is related to the assumption that all simulations and instructional measures used to support them were of similar implementation quality. The number of primary studies did not allow to differentiate between role of simulation features for each particular target skill and explore the relationship between the features in a greater detail.

To sum up, the large remaining (i.e., unexplained) heterogeneity of the effects requires caution when drawing conclusions about the effectiveness of specific types and combinations of scaffolding and technology.

Conclusion

There are hardly any study programs that would not aim to facilitate complex skills involving problem solving, diagnosing, communication, and collaboration. Simulations provide a wide range of practice opportunities and offer one of the most effective ways we know of designing learning environments in higher education. Simulation-based learning can start early in study programs, as it works well for beginners and advanced learners.

Although the analysis shows that social skills are not very enhanced, the acquisition of skills involving technical/manual performance can be facilitated a lot. The effect of simulation is greatly enhanced by the use of recent technologies. Higher levels of authenticity are related to greater effects while learning in both familiar and unfamiliar contexts. It is worth noting, however, that higher levels of

authenticity do not necessarily involve the use of recent technologies but rather more precise design of a simulation-based learning environment.

Scaffolding can additionally help, but the relative effect size compared with the effects of simulations is surprisingly small. However, rather than casting doubt on the relevance of scaffolding for simulation-based learning environments, we suggest trying to identify the most effective types, combinations, and sequences of scaffolding for learners with different prior knowledge and experience. Further research on scaffolding may investigate the optimal transitions of different types of scaffolding with increasing levels of complex skills. Including other kinds or instructional support in further research might provide important additional insights for designing effective simulations.

Note

ORCID iDs

Doris Holzberger 🆔 https://orcid.org/0000-0001-6535-289X
Tina Seidel 🆔 https://orcid.org/0000-0002-2578-1208

References

References marked with an asterisk indicate studies included in the meta-analysis.
*Abrahamson, S., & Denson, J. (1968). *A developmental study of medical training simulators for anesthesiologists: Final report* (Report No. BR-5-0917). University of Southern California. https://archive.org/stream/ERIC_ED019253/ERIC_ED019253_djvu.txt
*Adcock, A. B., Duggan, M. H., Watson, G. S., & Belfore, L. A. (2010). The impact of content area focus on the effectiveness of a web-based simulation. *British Journal of Educational Technology*, *41*(3), 388–402. https://doi.org/10.1111/j.1467-8535.2009.00947.x
*Ahad, S., Boehler, M., Schwind, C., & Hassan, I. (2013). The effect of model fidelity on colonoscopic skills acquisition: A randomized controlled study. *Journal of Surgical Education*, *70*(4), 522–527. https://doi.org/10.1016/j.jsurg.2013.02.010
*Ahlberg, G., Enochsson, L., Gallagher, A. G., Hedman, L., Hogman, C., McClusky, D. A., III, Ramel, S., Smith, D., & Arvidsson, D. (2007). Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *American Journal of Surgery*, *193*(6), 797–804. https://doi.org/10.1016/j.amjsurg.2006.06.050
*Ahlqvist, J., Nilsson, T., Hedman, L., Desser, T., Dev, P., Johansson, M., Youngblood, P., Cheng, R., & Gold, G. (2013). A randomized controlled trial on 2 simulation-based training methods in radiology effects on radiologic technology student skill in assessing image quality. *Simulation in Healthcare*, *8*(6), 382–387. https://doi.org/10.1097/SIH.0b013e3182a60a48
*Ahmad, R., Alhashmi, G., Ajlan, A., & Eldeek, B. (2015). Impact of high-fidelity transvaginal ultrasound simulation for radiology on residents' performance and

satisfaction. *Academic Radiology*, *22*(2), 234–239. https://doi.org/10.1016/j.acra.2014.09.006

*Ahn, H., & Kim, H.-Y. (2015). Implementation and outcome evaluation of high-fidelity simulation scenarios to integrate cognitive and psychomotor skills for Korean nursing students. *Nurse Education Today*, *35*(5), 706–711. https://doi.org/10.1016/j.nedt.2015.01.021

*Ainsworth, H., Gilchrist, M., Grant, C., Hewitt, C., Ford, S., Petrie, M., Torgerson, C., & Torgerson, D. (2011). Computer-based instruction for improving student nurses' general numeracy: Is it effective? Two randomised trials. *Educational Studies*, *38*(2), 1–13. https://doi.org/10.1080/03055698.2011.598668

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, *73*(3), 277–320. https://doi.org/10.3102/00346543073003277

*Alfred, M., & Chung, C. (2011). Design, development, and evaluation of a second generation interactive Simulator for Engineering Ethics Education (SEEE2). *Science and Engineering Ethics*, *18*(4), 689–697. https://doi.org/10.1007/s11948-011-9284-0

*Alinier, G., Hunt, B., Gordon, R., & Harwood, C. (2006). Effectiveness of intermediate-fidelity simulation training technology in undergraduate nursing education. *Journal of Advanced Nursing*, *54*(3), 359–369. https://doi.org/10.1111/j.1365-2648.2006.03810.x

Allen, J. A., Buffardi, L. C., & Hays, R. T. (1991). *The relationship of simulator fidelity to task and performance variables* (Report No. ARI-91-58). Army Research Institute for the Behavioral and Social Sciences. https://doi.org/10.21236/ADA238941

*Alyousef, S., Marwa, H., Alnojaidi, N., Lababidi, H., & Bashir, S. (2017). Cumulative evaluation data: Pediatric airway management simulation courses for pediatric residents. *Advances in Simulation*, *2*, Article 11. https://doi.org/10.1186/s41077-017-0044-3

*Andreatta, P., Chen, Y., Marsh, M., & Cho, K. (2010). Simulation-based training improves applied clinical placement of ultrasound-guided PICCs. *Supportive Care in Cancer*, *19*(4), 539–543. https://doi.org/10.1007/s00520-010-0849-2

*Andreatta, P., Hash, S., Klotz, J., Hauptman, J., Biddinger, B., & House, J. (2015). Performance-based comparison of neonatal intubation training outcomes simulator and live animal. *Advances in Neonatal Care*, *15*(1), 56–64. https://doi.org/10.1097/ANC.0000000000000130

*Andreatta, P., Hash, S., Klotz, J., Hauptman, J., Biddinger, B., & House, J. (2016). Retention curves for pediatric and neonatal intubation skills after simulation-based training. *Pediatric Emergency Care*, *32*(2), 71–76. https://doi.org/10.1097/PEC.0000000000000603

*Andreatta, P., Klotz, J., Madsen, J., Hurst, C., & Talbot, T. (2015). Outcomes from two forms of training for first-responder competency in cholinergic crisis management. *Military Medicine*, *180*(4), 468–474. https://doi.org/10.7205/MILMED-D-14-00290

*Aper, L., Reniers, J., Koole, S., Valcke, M., & Derese, A. (2012). Impact of three alternative consultation training formats on self-efficacy and consultation skills of medical students. *Medical Teacher*, *34*(7), 500–507. https://doi.org/10.3109/0142159X.2012.668627

*Araújo, S., Delaney, C., Seid, V., Imperiale, A., Bertoncini, A., Nahas, S., & Cecconello, I. (2014). Short-duration virtual reality simulation training positively

impacts performance during laparoscopic colectomy in animal model: Results of a single-blinded randomized trial—VR warm-up for laparoscopic colectomy. *Surgical Endoscopy*, *28*(9), 2547–2554. https://doi.org/10.1007/s00464-014-3500-3

*Atayee, R., Awdishu, L., & Namba, J. (2016). Using simulation to improve first-year pharmacy students' ability to identify medication errors involving the top 100 pre-scription medications. *American Journal of Pharmaceutical Education*, *80*(5), 86–96. https://doi.org/10.5688/ajpe80586

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.

*Auten, J., Ross, E., French, M., Li, I., Robinson, L., Brown, N., King, K., & Tanen, D. (2014). Low-fidelity hybrid sexual assault simulation training's effect on the comfort and competency of resident physicians. *Journal of Emergency Medicine*, *48*(3), 344–350. https://doi.org/10.1016/j.jemermed.2014.09.032

*Bachmann, C., Barzel, A., Roschlaub, S., Ehrhardt, M., & Scherer, M. (2013). Can a brief two-hour interdisciplinary communication skills training be successful in undergraduate medical education? *Patient Education and Counseling*, *93*(2), 298–305. https://doi.org/10.1016/j.pec.2013.05.019

*Banks, E., Chudnoff, S., Karmin, I., Wang, C., & Pardanani, S. (2007). Does a surgi-cal simulator improve resident operative performance of laparoscopic tubal ligation? *American Journal of Obstetrics & Gynecology*, *197*(5), 541.E1–545.E5. https://doi.org/10.1016/j.ajog.2007.07.028

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

Beaubien, J. M., & Baker, D. P. (2004). The use of simulation for training teamwork skills in health care: How low can you go? *BMJ Quality & Safety*, *13*(Suppl. 1), 51–56. https://doi.org/10.1136/qshc.2004.009845

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

*Bender, J., Kennally, K., Shields, R., & Overly, F. (2014). Does simulation booster impact retention of resuscitation procedural skills and teamwork? *Journal of Perinatology*, *34*(9), 664–668. https://doi.org/10.1038/jp.2014.72

*Bentley, S., Mudan, G., Strother, C., & Wong, N. (2015). Are live ultrasound models replaceable? Traditional versus simulated education module for FAST exam. *Western Journal of Emergency Medicine*, *16*(6), 818–822. https://doi.org/10.5811/westjem.2015.9.27276

*Biese, K., Moro-Sutherland, D., Furberg, R., Downing, B., Glickman, L., Murphy, A., Jackson, C., Snyder, G., & Hobgood, C. (2009). Using screen-based simulation to improve performance during pediatric resuscitation. *Academic Emergency Medicine*, *16*(2), 71–75. https://doi.org/10.1111/j.1553-2712.2009.00590.x

*Bjerrum, A., Hilberg, O., VanGog, T., Charles, P., & Eika, B. (2013). Effects of mod-elling examples in complex procedural skills training: A randomised study. *Medical Education*, *47*(9), 888–898. https://doi.org/10.1111/medu.12199

*Blackwood, J., Duff, J. P., Nettel-Aguirre, A., Djogovic, D., & Joynt, C. (2014). Does teaching crisis resource management skills improve resuscitation performance in pediatric residents? *Pediatric Critical Care Medicine*, *15*(4), 168–174. https://doi.org/10.1097/PCC.0000000000000100

*Boncyk, C., Schroeder, K., Anderson, B., & Galgon, R. (2016). Two methods for teaching basic upper airway sonography. *Journal of Clinical Anesthesia*, *31*(1), 166–172. https://doi.org/10.1016/j.jclinane.2016.01.040

*Bongers, P., Hove, P., Stassen, L., Dankelman, J., & Schreuder, H. W. (2014). A new virtual reality training module for laparoscopic surgical skills and equipment handling: Can multitasking be trained? A randomized controlled trial. *Journal of Surgical Education*, *72*(2), 184–191. https://doi.org/10.1016/j.jsurg.2014.09.004

*Bonnetain, E., Boucheix, J.-M., Hamet, M., & Freysz, M. (2010). Benefits of computer screen-based simulation in learning cardiac arrest procedures. *Medical Education*, *44*(7), 716–722. https://doi.org/10.1111/j.1365-2923.2010.03708.x

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Boshuizen, H. P. A., & Schmidt, H. (2008). The development of clinical reasoning expertise. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 113–122). Butterworth Heinemann.

*Broadbent, F. W., & Neehan, D. R. (1971). *An evaluation of simulation as an approach to assisting elementary teachers to identify children with learning disabilities and utilize ancillary personnel in initiating remediation programs within their classrooms: Final Report* (ED056425). ERIC. https://files.eric.ed.gov/fulltext/ED056425.pdf

*Brown, D., Miskovic, D., Tang, B., & Hanna, G. (2010). Impact of established skills in open surgery on the proficiency gain process for laparoscopic surgery. *Surgical Endoscopy*, *24*(6), 1420–1426. https://doi.org/10.1007/s00464-009-0792-9

*Brubacher, S. P., Powell, M., Skouteris, H., & Guadagno, B. (2015). The effects of e-simulation interview training on teachers' use of open-ended questions. *Child Abuse & Neglect*, *43*(1), 95–103. https://doi.org/10.1016/j.chiabu.2015.02.004

*Brunckhorst, O., Shahid, S., Aydın, A., Mcilhenny, C., Khan, S., Raza, S., Sahai, A., Brewin, J., Bello, F., Khan, M., Dasgupta, P., & Ahmed, K. (2015). Simulation-based ureteroscopy skills training curriculum with integration of technical and non-technical skills: A randomized controlled trial. *Surgical Endoscopy*, *29*(9), 2728–2735. https://doi.org/10.1007/s00464-014-3996-6

*Brydges, R., Nair, P., Ma, I., Shanks, D., & Hatala, R. (2012). Directed self-regulated learning versus instructor regulated learning in simulation training. *Medical Education*, *46*(7), 648–656. https://doi.org/10.1111/j.1365-2923.2012.04268.x

*Burton, K., Pendergrass, T., Byczkowski, T., Taylor, R., Moyer, M., Falcone, R., & Geis, G. (2011). Impact of simulation-based extracorporeal membrane oxygenation training in the simulation laboratory and clinical environment. *Simulation in Healthcare*, *6*(5), 284–291. https://doi.org/10.1097/SIH.0b013e31821dfcea

*Cannon, W., Garrett, W., Hunter, R., Sweeney, H., Eckhoff, D., Nicandri, G., Hutchinson, M., Johnson, D., Bisson, L., Bedi, A., Hill, J., Koh, J., & Reinig, K. (2014). Improving residency training in arthroscopic knee surgery with use of a virtual-reality simulator: A randomized blinded study. *Journal of Bone & Joint Surgery*, *96*(21), 1798–1806. https://doi.org/10.2106/JBJS.N.00058

Carroll, J. M., & Carrithers, C. (1984). Training wheels in a user interface. *Communications of the ACM*, *27*(8), 800–806. https://doi.org/10.1145/358198.358218

Carter, E., Schönbrodt, F., Gervais, W. M., & Hilgard, J. (2017). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

*Carvalho, I., Pais, V., Almeida, S., Ribeiro-Silva, R., Figueiredo-Braga, M., Teles, A., Castro-Vale, I., & Mota-Cardoso, R. (2011). Learning clinical communication skills: Outcomes of a program for professional practitioners. *Patient Education and Counseling*, *84*(1), 84–89. https://doi.org/10.1016/j.pec.2010.05.010

*Chao, C., Chalouhi, G., Bouhanna, P., Ville, Y., & Dommergues, M. (2015). Randomized clinical trial of virtual reality simulation training for transvaginal gynecologic ultrasound skills. *Journal of Ultrasound in Medicine*, *34*(9), 1663–1667. https://doi.org/10.7863/ultra.15.14.09063

*Chen, R., Grierson, L., & Norman, G. (2015). Evaluating the impact of high- and low-fidelity instruction in the development of auscultation skills. *Medical Education*, *49*(3), 276–285. https://doi.org/10.1111/medu.12653

*Cheng, H., Podolsky, D. J., Fisher, D. M., Wong, K. W., Lorenz, H. P., Khosla, R. K., Drake, J. M., & Forrest, C. R. (2018). Teaching palatoplasty using a high-fidelity cleft palate simulator. *Plastic and Reconstructive Surgery*, *141*(1), 91–98. https://doi.org/10.1097/PRS.0000000000003957

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2019). 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

*Chiu, M., Arab, A., Elliott, R., & Naik, V. (2011). An experiential teaching session on the anesthesia machine check improves resident performance. *Canadian Journal of Anaesthesia/Journal canadien d'anesthésie*, *59*(3), 280–287. https://doi.org/10.1007/s12630-011-9649-5

*Chung, C., Cooper, S., Cant, R., Connell, C., Mckay, A., Kinsman, L., Gazula, S., Boyle, J., Cameron, A., Cash, P., Evans, L., Kim, J., Masud, R., McInnes, D., Norman, L., Penz, E., Rotter, T., Tanti, E., & Breakspear, T. (2018). The educational impact of web-based and face-to-face patient deterioration simulation programs: An interventional trial. *Nurse Education Today*, *64*(1), 93–98. https://doi.org/10.1016/j.nedt.2018.01.037

*Chung, G. K. W. K., Gyllenhammer, R. G., & Baker, E. L. (2011). *The effects of practicing with a virtual ultrasound trainer on FAST window identification, acquisition, and diagnosis* (CRESST Report 787). National Center for Research on Evaluation, Standards, and Student Testing. http://cresst.org/wp-content/uploads/R787.pdf

*Clayton, J., Butow, P., Waters, A., Laidsaar-Powell, R., O'Brien, A., Boyle, F., Back, A., Arnold, R., Tulsky, J., & Tattersall, M. (2012). Evaluation of a novel individualised communication-skills training intervention to improve doctors' confidence and skills in end-of-life communication. *Palliative Medicine*, *27*(3), 236–243. https://doi.org/10.1177/0269216312449683

Cook, D. A. (2014). How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education. *Medical Education*, *48*(8), 750–760. https://doi.org/10.1111/medu.12473

Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J., & Hatala, R. (2013). Technology-enhanced simulation to assess health professionals: A systematic review of validity evidence, research methods, and reporting quality. *Academic Medicine*, *88*(6), 872–883. https://doi.org/10.1097/ACM.0b013e31828ffdcf

*Cuisinier, A., Schilte, C., Declety, P., Picard, J., Berger, K., Bouzat, P., Falcon, D., Bosson, J.-L., Payen, J.-F., & Albaladejo, P. (2015). A major trauma course based on posters, audio-guides and simulation improves the management skills of medical

students: Evaluation via medical simulator. *Anaesthesia Critical Care & Pain Medicine*, *34*(6), 339–344. https://doi.org/10.1016/j.accpm.2015.06.009

*Damle, L. F., Tefera, E., McAfee, J., Loyd, M. K., Jackson, A. M., Auguste, T. C., & Gomez-Lobo, V. (2015). Pediatric and Adolescent Gynecology Education through Simulation (PAGES): Development and evaluation of a simulation curriculum. *Journal of Pediatric & Adolescent Gynecology*, *28*(3), 186–191. https://doi.org/10.1016/j.jpag.2014.07.008

D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., & Haertel, G. (2014). *Simulations for STEM learning: Systematic review and meta-analysis*. SRI International. https://www.sri.com/publication/simulations-for-stem-learning-systematic-review-and-meta-analysis-full-report/

*Dankbaar, M., Alsma, J., Jansen, E. E., Merrienboer, J. J., Saase, J. L., & Schuit, S. (2016). An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Advances in Health Sciences Education*, *21*(3), 505–521. https://doi.org/10.1007/s10459-015-9641-x

Davidsson, P., & Verhagen, H. (2017). Types of simulation. In B. Edmonds & R. Meyer (Eds.), *Simulating social complexity. Understanding complex systems* (pp. 23–37). Springer. https://doi.org/10.1007/978-3-319-66948-9_3

*DeWaay, D. J., McEvoy, M. D., Alexander, L. A., Kern, D. H., & Nietert, P. J. (2014). Simulation curriculum can improve medical student assessment and management of acute coronary syndrome during a clinical practice exam. *American Journal of the Medical Sciences*, *347*(6), 452–456. https://doi.org/10.1097/MAJ.0b013e3182a562d7

*Ditton-Phare, P., Sandhu, H., Kelly, B., Kissane, D., & Loughland, C. (2016). Pilot evaluation of a communication skills training program for psychiatry residents using standardized patient assessment. *Academic Psychiatry*, *40*(5), 768–775. https://doi.org/10.1007/s40596-016-0560-9

*Djukic, M., Adams, J., Fulmer, T., Szyld, D., Lee, S., Oh, S.-Y., & Triola, M. (2015). E-learning with virtual teammates: A novel approach to interprofessional education. *Journal of Interprofessional Care*, *29*(5), 476–482. https://doi.org/10.3109/13561820.2015.1030068

Dochy, F., Segers, M., van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*(5), 533–568. https://doi.org/10.1016/S0959-4752(02)00025-7

*Douglas, S., Andrade, A., Boyd, S., Leslie, M., Webb, L., Davis, L., Fraine, M., Frazer, N., Hargraves, R., & Bickman, L. (2016). Communication training improves patient-centered provider behavior and screening for soldiers' mental health concerns. *Patient Education and Counseling*, *99*(7), 1203–1212. https://doi.org/10.1016/j.pec.2016.01.018

*Dubovi, I., Levy, S., & Dagan, E. (2017). Now I know how! The learning process of medication administration among nursing students with non-immersive desktop virtual reality simulation. *Computers & Education*, *113*, 16–27. https://doi.org/10.1016/j.compedu.2017.05.009

*Dumont, T., Hakim, J., Black, A., & Fleming, N. (2016). Does an advanced pelvic simulation curriculum improve resident performance on a pediatric and adolescent gynecology focused objective structured clinical examination? A cohort study. *Journal of Pediatric & Adolescent Gynecology*, *29*(3), 276–279. https://doi.org/10.1016/j.jpag.2015.10.015

*Durmaz, A., Sarıkaya, A., Cakan, E., & Cakir, S. (2012). Effect of screen-based computer simulation on knowledge and skill in nursing students' learning of

preoperative and postoperative care management: A randomized controlled study. *Computers, Informatics, Nursing: CIN*, *30*(4), 196–203. https://doi.org/10.1097/NCN.0b013e3182419134

*Eghbalibabadi, M., & Ashouri, E. (2014). Comparison of the effects of two teaching methods on the nursing students' performance in measurement of blood pressure. *Iranian Journal of Nursing and Midwifery Research*, *19*(4), 381–384. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4145493/

*Errek, H., & Randolph, D. (1982). Effects of discussion and role-play activities in the acquisition of consultant interview skills. *Journal of Counseling Psychology*, *29*(3), 304–308. https://doi.org/10.1037/0022-0167.29.3.304

*Etezadi, F., Najafi, A., Pourfakhr, P., Moharari, R., Khajavi, M., Imani, F., & Barzin, G. (2016). An assessment of intubation skill training in novice anesthesiology residents of Tehran University of Medical Sciences with the use of mannequins. *Anesthesiology and Pain Medicine*, *6*(6), Article e39184. https://doi.org/10.5812/aapm.39184

*Evans, K., Daines, W., Tsui, J., Strehlow, M., Maggio, P., & Shieh, L. (2014). Septris: A novel, mobile, online, simulation game that improves sepsis recognition and management. *Academic Medicine*, *90*(2), 180–184. https://doi.org/10.1097/ACM.0000000000000611

*Fallucco, E. M., Conlon, M. K., Gale, G. W., Constantino, J. N., & Glowinski, A. (2012). Use of a standardized patient paradigm to enhance proficiency in risk assessment for adolescent depression and suicide. *Journal of Adolescent Health*, *51*(1), 66–72. https://doi.org/10.1016/j.jadohealth.2011.12.026

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*(2), 28–45. https://doi.org/10.14786/flr.v2i2.96

*Fisher, N., Eisen, L. A., Bayya, J. V., Dulu, A., Bernstein, P. S., Merkatz, I. R., & Goffman, D. (2011). Improved performance of maternal-fetal medicine staff after maternal cardiac arrest simulation-based training. *American Journal of Obstetrics & Gynecology*, *205*(3), 239.E1–239.E5. https://doi.org/10.1016/j.ajog.2011.06.012

*Fleming, M., Olsen, D., Stathes, H., Boteler, L., Grossberg, P., Pfeifer, J., Schiro, S., Banning, J., & Skochelak, S. (2009). Virtual reality skills training for health care professionals in alcohol screening and brief intervention. *Journal of the American Board of Family Medicine*, *22*(4), 387–398. https://doi.org/10.3122/jabfm.2009.04.080208

*Foster, A., Chaudhary, N., Murphy, J., Lok, B., Waller, J., & Buckley, P. (2014). The use of simulation to teach suicide risk assessment to health profession trainees: Rationale, methodology, and a proof of concept demonstration with a virtual patient. *Academic Psychiatry*, *39*(6), 620–629. https://doi.org/10.1007/s40596-014-0185-9

Funke, J. (Ed.). (2006). *Denken und Problemlösen* (pp. 375−446). Hogrefe.

*Gable, B., Gardner, A., Celik, D., Bhalla, M., & Ahmed, R. (2014). Improving bariatric patient transport and care with simulation. *Western Journal of Emergency Medicine*, *15*(2), 199–204. https://doi.org/10.5811/westjem.2013.12.18855

Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology*, *45*(6), 1097–1114. https://doi.org/10.1111/bjet.12188

*Girod, M., & Girod, G. (2006). Exploring the efficacy of the Cook School District simulation. *Journal of Teacher Education*, *57*(5), 481−497. https://doi.org/10.1177/0022487106293742

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

*Grover, S., Garg, A., Scaffidi, M., Yu, J., Plener, I., Yong, E., Cino, M., Grantcharov, T., & Walsh, C. (2015). Impact of a simulation training curriculum on technical and nontechnical skills in colonoscopy: A randomized trial. *Gastrointestinal Endoscopy*, *82*(6), 1072–1079. https://doi.org/10.1016/j.gie.2015.04.008

*Grugnetti, A., Bagnasco, A., & Sasso, L. (2013). Effectiveness of a clinical skills workshop for drug-dosage calculation in a nursing program. *Nurse Education Today*, *34*(4), 619–624. https://doi.org/10.1016/j.nedt.2013.05.021

*Haak, R., Rosenbohm, J., Koerfer, A., Obliers, R., & Wicht, M. (2008). The effect of undergraduate education in communication skills: A randomised controlled clinical trial. *European Journal of Dental Education*, *12*(4), 213–218. https://doi.org/10.1111/j.1600-0579.2008.00521.x

*Hamilton, E., Scott, D., Kapoor, A., Nwariaku, F., Bergen, P., Rege, R., Tesfay, S., & Jones, D. B. (2002). Improving operative performance using a laparoscopic hernia simulator. *American Journal of Surgery*, *182*(6), 725–728. https://doi.org/10.1016/S0002-9610(01)00800-5

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

*Harris, M., Pittiglio, L., Newton, S., & Moore, G. (2014). Using simulation to improve the medication administration skills of undergraduate nursing students. *Nursing Education Perspectives*, *35*(1), 26–29. https://doi.org/10.5480/11-552.1

Hattie, J. (2003). *Formative and summative interpretations of assessment information*. http://assessment.tki.org.nz/content/download/6076/61425/version/1/file/formative-and-summative-assessment-%282003%29.pdf

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

*Hebbar, K., Cunningham, C., McCracken, C., Kamat, P., & Fortenberry, J. (2014). Simulation-based paediatric intensive care unit central venous line maintenance bundle training. *Intensive and Critical Care Nursing*, *31*(1), 44–50. https://doi.org/10.1016/j.iccn.2014.10.003

*Hecimovich, M., & Volet, S. (2014). Simulated learning in musculoskeletal assessment and rehabilitation education: Comparing the effect of a simulation-based learning activity with a peer-based learning activity. *BMC Medical Education*, *14*(1), Article 253. https://doi.org/10.1186/s12909-014-0253-6

Hegland, P. A., Aarlie, H., Strømme, H., & Jamtvedt, G. (2017). Simulation-based training for nurses: Systematic review and meta-analysis. *Nurse Education Today*, *54*(1), 6–20. https://doi.org/10.1016/j.nedt.2017.04.004

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M., Ufer, S., Schmidmaier, R., Neuhaus, B., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Girwidz, R., & Fischer, F. (2019). Facilitating diagnostic competences in simulations: A conceptual framework and a research agenda for medical and teacher education. *Frontline Learning Research*, *7*(4), 1–24. https://doi.org/10.14786/flr.v7i4.384

*Helder, M. K., Rowse, P., Ruparel, R., Li, Z., Farley, D., Joyce, L., & Stulak, J. (2015). Basic cardiac surgery skills on sale for $22.50: An aortic anastomosis simulation curriculum. *Annals of Thoracic Surgery*, *101*(1), 316–322. https://doi.org/10.1016/j.athoracsur.2015.08.005

*Henney, M., & Boysen, V. (1979). The effect of computer simulation training on ability to administer an informal reading inventory. *Journal of Educational Research*, *72*(5), 265–270. https://doi.org/10.1080/00220671.1979.10885168

Henninger, M., & Mandl, H. (2000). Vom Wissen zum Handeln—ein Ansatz zur Förderung kommunikativen Handelns [From knowledge to action—An approach for fostering communicative behavior]. In H. Mandl & J. Gerstenmaier (Eds.), *Die Kluft zwischen Wissen und Handeln* (pp. 197–219). Hogrefe.

*Heskin, L., Mansour, E., Lane, B., Kavanagh, D., Dicker, P., Ryan, D., Gildea-Byrne, K., Pawlikowska, T., Tierney, S., & Traynor, O. (2015). The impact of a surgical boot camp on early acquisition of technical and nontechnical skills by novice surgical trainees. *American Journal of Surgery*, *210*(3), 570–577. https://doi.org/10.1016/j.amjsurg.2014.12.046

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hmelo-Silver, C. E., Dunkan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

*Hobgood, C., Harward, D., Newton, K., & Davis, W. (2005). The educational intervention "GRIEV_ING" improves the death notification skills of residents. *Academic Emergency Medicine*, *12*(4), 296–301. https://doi.org/10.1197/j.aem.2004.12.008

*Johnson, D., Corrigan, T., Gulickson, G., Holshouser, E., & Johnson, S. (2012). The effects of a human patient simulator vs. a CD-ROM on performance. *Military Medicine*, *177*(10), 1131–1135. https://doi.org/10.7205/MILMED-D-12-00179

*Johnson, T., Lyons, R., Kopper, R., Johnsen, K., Lok, B., & Cendan, J. (2014). Virtual patient simulations and optimal social learning context: A replication of an aptitude-treatment interaction effect. *Medical Teacher*, *36*(6), 486–494. https://doi.org/10.3109/0142159X.2014.890702

*Jones, J., Staub, J., Seymore, A., & Scott, L. A. (2014). Securing the second front: Achieving first receiver safety and security through competency-based tools. *Prehospital and Disaster Medicine*, *29*(6), 643–647. https://doi.org/10.1017/s1049023x14001058

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

*Kash, K., Leas, B., Clough, J., Dodick, D., Capobianco, D., Nash, D., & Bance, L. (2009). ACGME competencies in neurology: Web-based objective simulated computerized clinical encounters. *Neurology*, *72*(10), 893–898. https://doi.org/10.1212/01.wnl.0000344164.98457.bf

*Keleekai-Brapoh, N., Schuster, C., Murray, C., King, M., Stahl, B., Labrozzi, L., Gallucci, S., Leclair, M., & Glover, K. (2016). Improving nurses' peripheral intravenous catheter insertion knowledge, confidence, and skills using a simulation-based blended learning program: A randomized trial. *Simulation in Healthcare*, *11*(6), 376–384. https://doi.org/10.1097/SIH.0000000000000186

*Khatib, M., Hald, N., Brenton, H., Barakat, M. F., Sarker, S., Standfield, N., Ziprin, P., & Bello, F. (2014). Validation of open inguinal hernia repair simulation model— A randomised controlled educational trial. *American Journal of Surgery*, *208*(2), 295–301. https://doi.org/10.1016/j.amjsurg.2013.12.007

*Kiersma, M. E., Darbishire, P. L., Plake, K. S., Oswald, C. A., & Walters, B. M. (2009). Laboratory session to improve first-year pharmacy students' knowledge and confidence concerning the prevention of medication errors. *American Journal of Pharmaceutical Education*, *73*(6), Article 99. https://doi.org/10.5688/aj730699

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

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

*Koparan, T., & Yılmaz, G. (2015). The effect of simulation-based learning on prospective teachers' inference skills in teaching probability. *Universal Journal of Educational Research*, *3*(11), 775–786. https://doi.org/10.13189/ujer.2015.031101

*Kumar, D. D., & Sherwood, R. D. (2007). Effect of a problem-based simulation on the conceptual understanding of undergraduate science education students. *Journal of Science Education and Technology*, *16*(3), 239–246. https://doi.org/10.1007/s10956-007-9049-3

*Kwon, S., Hong, S. H., Kim, E., Park, H., You, Y., & Kim, Y.-H. (2015). The efficacy of lumbosacral spine phantom to improve resident proficiency in performing ultrasound-guided spinal procedure. *Pain Medicine*, *16*(12), 2284–2291. https://doi.org/10.1111/pme.12870

*Lavelle, M., Attoe, C., Tritschler, C., & Cross, S. (2017). Managing medical emergencies in mental health settings using an interprofessional in situ simulation training programme: A mixed methods evaluation study. *Nurse Education Today*, *59*(1), 103–109. https://doi.org/10.1016/j.nedt.2017.09.009

*Lee Chin, K., Yap, C., Lee, W. L., & Soh, Y. (2014). Comparing effectiveness of high-fidelity human patient simulation vs case-based learning in pharmacy education. *American Journal of Pharmaceutical Education*, *78*(8), Article 153. https://doi.org/10.5688/ajpe788153

*Lehmann, R., Bosse, H. M., Simon, A., Nikendei, C., & Huwendiek, S. (2013). An innovative blended learning approach using virtual patients as preparation for skills laboratory training: Perceptions of students and tutors. *BMC Medical Education*, *13*, Article 23. https://doi.org/10.1186/1472-6920-13-23

*Liao, W.-C., Leung, J., Wang, H.-P., Chang, W.-H., Chu, C.-H., Lin, J.-T., Wilson, R., Lim, B., & Leung, F. (2013). Coached practice using ERCP mechanical simulator improves trainees' ERCP performance: A randomized controlled trial. *Endoscopy*, *45*(10), 799–805. https://doi.org/10.1055/s-0033-1344224

*Liaw, S., Chan, S., Chen, F. G., Hooi, S., & Siau, C. (2014). Comparison of virtual patient simulation with mannequin-based simulation for improving clinical performances in assessing and managing clinical deterioration: Randomized controlled trial. *Journal of Medical Internet Research*, *16*(9), Article e214. https://doi.org/10.2196/jmir.3322

*Liaw, S., Lai, F. W., Chan, S., Ho, J., Mordiffi, S., Ang, S., Goh, P.-S., & Ang, E. (2015). Designing and evaluating an interactive multimedia Web-based simulation for developing nurses' competencies in acute nursing care: Randomized controlled

trial. *Journal of Medical Internet Research*, *17*(1), Article e5. https://doi.org/10.2196/jmir.3853

*Liaw, S., Rethans, J.-J., Scherpbier, A., & Klainin-Yobas, P. (2011). Rescuing a Patient in Deteriorating Situations (RAPIDS): A simulation-based educational program on recognizing, responding and reporting of physiological signs of deterioration. *Resuscitation*, *82*(9), 1224–1230. https://doi.org/10.1016/j.resuscitation.2011.04.014

*Madan, A. K., Caruso, B., Lopes, J. E., & Gracely, E. J. (1998). Comparison of simulated patient and didactic methods of teaching HIV risk assessment to medical residents. *American Journal of Preventive Medicine*, *15*(2), 114–119. https://doi.org/10.1016/s0749-3797(98)00026-9

*Maertens, H., Aggarwal, R., Moreels, N., Vermassen, F., & Van Herzeele, I. (2017). A Proficiency Based Stepwise Endovascular Curricular Training (PROSPECT) program enhances operative performance in real life: A randomised controlled trial. *European Journal of Vascular & Endovascular Surgery*, *54*(3), 387–396. https://doi.org/10.1016/j.ejvs.2017.06.011

Mann, K., Gordon, J., & MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education*, *14*(4), 595–621. https://doi.org/10.1007/s10459-007-9090-2

*Martin, K., Patterson, D., Phisitkul, P., Cameron, K., Femino, J., & Amendola, A. (2015). Ankle arthroscopy simulation improves basic skills, anatomic recognition, and proficiency during diagnostic examination of residents in training. *Foot & Ankle International*, *36*(7), 827–835. https://doi.org/10.1177/1071100715576369

*Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G., & Koedinger, K. (2013). Studying the effect of a competitive game show in a learning by teaching environment. *International Journal of Artificial Intelligence in Education*, *23*(1–4), 1–21. https://doi.org/10.1007/s40593-013-0009-1

Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2nd ed.). Freeman.

*McIntosh, K., Gregor, J., & Khanna, N. (2014). Computer-based virtual reality colonoscopy simulation improves patient-based colonoscopy performance. *Canadian Journal of Gastroenterology and Hepatology*, *28*(4), 203–206. https://doi.org/10.1155/2014/804367

*Nelissen, E., Ersdal, H., Mduma, E., Evjen-Olsen, B., Broerse, J., Roosmalen, J., & Stekelenburg, J. (2015). Helping mothers survive bleeding after birth: Retention of knowledge, skills, and confidence nine months after obstetric simulation-based training. *BMC Pregnancy and Childbirth*, *15*, Article 190. https://doi.org/10.1186/s12884-015-0612-2

*Newell, M., & Newell, T. S. (2018). Analyzing the effect of consultation training on the development of consultation competence. *Contemporary School Psychology*, *22*(1), 40–50. https://doi.org/10.1007/s40688-017-0151-0

*Ogan, K., Jacomides, L., Shulman, M., Roehrborn, C., Cadeddu, J., & Pearle, M. (2004). Virtual ureteroscopy predicts ureteroscopic proficiency of medical students on a cadaver. *Journal of Urology*, *172*(2), 667–671. https://doi.org/10.1097/01.ju.0000131631.60022.d9

*Ortner, C., Richebé, P., Bollag, L., Ross, B. K., & Landau, R. (2014). Repeated simulation-based training for performing general anesthesia for emergency cesarean delivery: Long-term retention and recurring mistakes. *International Journal of Obstetric Anesthesia*, *23*(4), 341–347. https://doi.org/10.1016/j.ijoa.2014.04.008

*O'Sullivan, O., Iohom, G., O'Donnell, B., & Shorten, G. (2014). The effect of simulation-based training on initial performance of ultrasound-guided axillary brachial plexus blockade in a clinical setting: A pilot study. *BMC Anesthesiology*, *14*(1), Article 110. https://doi.org/10.1186/1471-2253-14-110

*Overbaugh, R. (1995). The efficacy of interactive video for teaching basic classroom management skills to pre-service teachers. *Computers in Human Behavior*, *11*(3–4), 511–527. https://doi.org/10.1016/0747-5632(95)80014-Y

*Pantziaras, I., Fors, U., & Ekblad, S. (2015). Training with virtual patients in transcultural psychiatry: Do the learners actually learn? *Journal of Medical Internet Research*, *17*(2), Article e46. https://doi.org/10.2196/jmir.3497

*Passman, M. A., Fleser, P. S., Dattilo, J. B., Guzman, R. J., & Naslund, T. C. (2007). Should simulator-based endovascular training be integrated into general surgery residency programs? *American Journal of Surgery*, *194*(2), 212–219. https://doi.org/10.1016/j.amjsurg.2006.11.029

*Popadiuk, C., Pottle, M., & Curran, V. (2002). Teaching digital rectal examinations to medical students: An evaluation study of teaching methods. *Academic Medicine*, *77*(11), 1140–1146. https://doi.org/10.1097/00001888-200211000-00017

P21: Partnership for 21st Century Learning. (2019). *P21 framework for 21st century learning definitions*. Battelle for Kids. http://static.battelleforkids.org/documents/p21/P21_Framework_DefinitionsBFK.pdf

*Pucher, P., Aggarwal, R., Qurashi, M., Singh, P., & Darzi, A. (2014). Randomized clinical trial of the impact of surgical ward-care checklists on postoperative care in a simulated environment. *British Journal of Surgery*, *101*(13), 1666–1673. https://doi.org/10.1002/bjs.9654

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(3), 337–386. https://doi.org/10.1207/s15327809jls1303_4

*Rajan, S., Khanna, A., Argalious, M., Kimatian, S., Mascha, E., Makarova, N., Nada, E., Elsharkawy, H., Firoozbakhsh, F., & Avitsian, R. (2016). Comparison of 2 resident learning tools—Interactive screen-based simulated case scenarios versus problem-based learning discussions: A prospective quasi-crossover cohort study. *Journal of Clinical Anesthesia*, *28*, 4–11. https://doi.org/10.1016/j.jclinane.2015.08.003

*Randell, T., Hall, M., Bizo, L., & Remington, B. (2007). DTkid: Interactive simulation software for training tutors of children with autism. *Journal of Autism and Developmental Disorders*, *37*(4), 637–647. https://doi.org/10.1007/s10803-006-0193-z

Raven, J. (2000). Psychometrics, cognitive ability, and occupational performance. *Review of Psychology*, *7*(1–2), 51–74.

*Reis, S., Sagi, D., Eisenberg, O., Kuchnir, Y., Azuri, J., Shalev, V., & Ziv, A. (2013). The impact of residents' training in electronic medical record (EMR) use on their competence: Report of a pragmatic trial. *Patient Education and Counseling*, *93*(3), 515–521. https://doi.org/10.1016/j.pec.2013.08.007

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

Rogers, P. L. (2001). Traditions to transformations: The forced evolution of higher education. *AACE Journal*, *9*(1), 47–60.

*Ross, W., Pollman, W., Perry, D., Welty, J., & Jones, K. (2001). Interactive video negotiator training: A preliminary evaluation of the McGill negotiation simulator. *Simulation & Gaming*, *32*(4), 451–468. https://doi.org/10.1177/104687810103200402

*Roter, D. L., Cole, K. A., Kern, D. E., Barker, L. R., & Grayson, M. (1990). An evaluation of residency training in interviewing skills and the psychosocial domain of medical practice. *Journal of General Internal Medicine*, *5*(4), 347–354. https://doi.org/10.1007/BF02600404

*Roter, D. L., Edelman, E., Larson, S., McNellis, R., Erby, L., Massa, M., Rackover, M., & McInerney, J. (2012). Effects of online genetics education on physician assistant interviewing skills. *JAAPA: Official Journal of the American Academy of Physician Assistants*, *25*(8), 36–38. https://doi.org/10.1097/01720610-201208000-00007

*Saraf, S., Bayya, J., Weedon, J., Minkoff, H., & Fisher, N. (2014). The relationship of praise/criticism to learning during obstetrical simulation: A randomized clinical trial. *Journal of Perinatal Medicine*, *42*(4), 1–8. https://doi.org/10.1515/jpm-2013-0247

*Sauter, T., Hautz, W., Hostettler, S., Brodmann Maeder, M., Martinolli, L., Lehmann, B., Exadaktylos, A., & Haider, D. (2016). Interprofessional and interdisciplinary simulation-based training leads to safe sedation procedures in the emergency department. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *24*, Article 97. https://doi.org/10.1186/s13049-016-0291-7

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, *5*(3), 205–221. https://doi.org/10.1007/BF01323044

Schmidt, H. G., Loyens, S. M. M., van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 91–97. https://doi.org/10.1080/00461520701263350

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education*, *41*(12), 1133–1139. https://doi.org/10.1111/j.1365-2923.2007.02915.x

*Scott, L., Swartzentruber, D., Davis, C., Maddux, P., Schnellmann, J., & Wahlquist, A. (2013). Competency in chaos: Lifesaving performance of care providers utilizing a competency-based, multi-actor emergency preparedness training curriculum. *Prehospital and Disaster Medicine*, *28*(4), 1–12. https://doi.org/10.1017/S1049023X13000368

*Sheakley, M., Gilbert, G., Leighton, K., Hall, M., Callender, D., & Pederson, D. (2016). A brief simulation intervention increasing basic science and clinical knowledge. *Medical Education Online*, *21*(1). https://doi.org/10.3402/meo.v21.30744

Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching*, *40*(1), 6–33. https://doi.org/10.1002/tea.10058

*Sibbald, M., McKinney, J., Cavalcanti, R. B., Yu, E. H., Wood, D. A., Nair, P. S., Eva, K. W., & Hatala, R. (2013). Cardiac examination and the effect of dual-processing instruction in a cardiopulmonary simulator. *Advances in Health Sciences Education*, *18*(3), 497–508. https://doi.org/10.1007/s10459-012-9388-6

*Skinner, A., Freeman, R., & Sheehan, F. (2016). Quantitative feedback facilitates acquisition of skills in focused cardiac ultrasound. *Simulation in Healthcare*, *11*(2), 134–138. https://doi.org/10.1097/SIH.0000000000000132

*Solomon, D., Laird-Fick, H., Keefe, C., Thompson, M., & Noel, M. (2004). Using a formative simulated patient exercise for curriculum evaluation. *BMC Medical Education*, *4*, Article 8. https://doi.org/10.1186/1472-6920-4-8

*Sorensen, J., Van der Vleuten, C., Rosthøj, S., Østergaard, D., LeBlanc, V., Johansen, M., Ekelund, K., Starkopf, L., Lindschou, J., Gluud, C., Weikop, P., & Ottesen, B. (2015). Simulation-based multiprofessional obstetric anaesthesia training conducted *in situ* versus off-site leads to similar individual and team outcomes: A randomised educational trial. *British Medical Journal Open*, *5*(10), Article e008344. https://doi.org/10.1136/bmjopen-2015-008344corr1

*Sperl-Hillen, J., O'Connor, P., Ekstrom, H., Rush, W., Asche, S., Fernandes, O., Apana, D., Amundson, G., Johnson, P., & Curran, D. (2014). Educating resident physicians using virtual case-based simulation improves diabetes management: A randomized controlled trial. *Academic Medicine*, *89*(12), 1664–1673. https://doi.org/10.1097/ACM.0000000000000406

*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

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055. https://doi.org/10.1016/S0895-4356(01)00377-8

Tabak, I., & Kyza, E. (2018). Research on scaffolding in the learning sciences: A methodological perspective. In F. Fischer, C. Hmelo-Silver, S. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 191–200). Routledge.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, *81*(1), 4–28. https://doi.org/10.3102/0034654310393361

Tanner-Smith, E., Tipton, E., & Polanin, J. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. https://doi.org/10.1007/s40865-016-0026-5

*Ten Eyck, R., Tews, M., Ballester, J., & Hamilton, G. (2010). Improved fourth-year medical student clinical decision-making performance as a resuscitation team leader after a simulation-based curriculum. *Simulation in Healthcare*, *5*(3), 139–145. https://doi.org/10.1097/SIH.0b013e3181cca544

*Tobin, C., Clark, C., Mcevoy, M., Reves, J. G., Schaefer, J., Wolf, B., & Reeves, S. (2013). An approach to moderate sedation simulation training. *Simulation in Healthcare*, *8*(2), 114–123. https://doi.org/10.1097/SIH.0b013e3182786209

Van, de, Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

Van Lehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*, 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

Veritas Health Innovation. (2019). *Covidence systematic review software* (online version) [Computer software]. https://www.covidence.org

*Wenk, M., Waurick, R., Schotes, D., Wenk, M., Gerdes, C., Van Aken, H., & Pöpping, D. (2008). Simulation-based medical education is no better than problem-based discussions and induces misjudgment in self-assessment. *Advances in Health Sciences Education*, *14*(2), 159–171. https://doi.org/10.1007/s10459-008-9098-2

*Wheatley, W. J., Hornaday, R. W., & Hunt, T. G. (1988). Developing strategic management goal-setting skills. *Simulation & Games*, *19*(2), 173–185. https://doi.org/10.1177/104687818801900205

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

Wu, Y., & Anderson, O. R. (2015). Technology-enhanced STEM (science, technology, engineering, and mathematics) education. *Journal of Computers in Education*, *2*(3), 245–249. https://doi.org/10.1007/s40692-015-0041-2

*Yeh, M.-L., & Chen, H.-H. (2005). Effects of an educational program with interactive videodisc systems in improving critical thinking dispositions for RN-BSN students in Taiwan. *International Journal of Nursing Studies*, *42*(3), 333–340. https://doi.org/10.1016/j.ijnurstu.2004.06.008

Authors

OLGA CHERNIKOVA currently holds a PhD in learning sciences from Ludwig-Maximilians-Universität in Munich, Leopoldstrasse 13, Munich 80802, Germany; email: *o.chernikova@psy.lmu.de*. She holds a position of a research fellow at the chair of Educational Psychology and Educational Sciences at LMU Munich. Her research interests deal largely with use of digital media and instructional support in teacher education.

NICOLE HEITZMANN holds a doctoral degree in educational sciences. She is a postdoc research fellow at the Munich Center of the Learning Sciences. She is affiliated with the Department of Psychology and the Institute of Medical Education at Ludwig-Maximilians-Universität in Munich, Leopoldstrasse 13, Munich 80802, Germany; email: *nicole.heitzmann@psy.lmu.de*.

MATTHIAS STADLER is an assistant professor at the chair of Educational Psychology and Educational Sciences, Department of Psychology, at Ludwig-Maximilians-Universität in Munich, Leopoldstrasse 13, Munich 80802, Germany; email: *matthias .stadler@psy.lmu.de*. His research interests lie in the fields of educational psychology and assessment with a focus on computer-based assessment and simulations.

DORIS HOLZBERGER is an associate professor of research on learning and instruction at the Centre for International Student Assessment, Technical University of Munich, Arcisstrasse 21, Munich 80333, Germany; email: *doris.holzberger@tum.de*. She conducts empirical research into education with a focus on learning and instruction, especially in the field of teacher characteristics and instructional quality.

TINA SEIDEL is a full professor of educational psychology at the Technical University of Munich, Arcisstrasse 21, Munich 80333, Germany; email: *tina.seidel@tum.de*. Her research interests are focused on teaching and teacher research, particularly in the fields of professional vision, teacher pedagogical–psychological knowledge, and teacher–student interactions in classrooms.

FRANK FISCHER is a full professor of educational psychology and educational sciences, Department of Psychology, at Ludwig-Maximilians-Universität in Munich, Germany, and is Director of the Munich Center of the Learning Sciences, Leopoldstrasse 13, Munich 80802, Germany; email: *frank.fischer@psy.lmu.de*.