Association between Clinical Simulation Design Features and Novice Healthcare Professionals’ Cognitive Load: A Systematic Review and Meta-Analysis

Alexandra Lapierre\textsuperscript{1,2}, Caroline Arbour\textsuperscript{1,2}, Marc-André Maheu-Cadotte\textsuperscript{1}, Billy Vinette\textsuperscript{1}, Guillaume Fontaine\textsuperscript{3}, and Patrick Lavoie\textsuperscript{1}

Abstract

Background. Clinical simulations are complex educational interventions characterized by several design features, which have the potential to influence cognitive load, that is, the mental effort required to assimilate new information and learn. This systematic review and meta-analysis explored the associations between simulation design features and cognitive load in novice healthcare professionals.

Methods. Based on the Joanna Briggs Institute methodology, a search was performed in five databases for quantitative studies in which the cognitive load of novice healthcare professionals was measured during or after a simulation activity. Each clinical simulation was coded to describe its design features. Univariate and multivariate mixed model analyses were performed to explore the associations between simulation design features and cognitive load.

\textsuperscript{1}Faculty of Nursing, Université de Montréal, Montreal, QC, Canada
\textsuperscript{2}Research centre, Montreal Sacré-Cœur Hospital, Montreal, QC, Canada
\textsuperscript{3}Faculty of Medicine, University of Ottawa, Ottawa, QC, Canada

Corresponding Author:
Alexandra Lapierre, Faculty of Nursing, Université de Montréal, C. P. Succ. Centre-Ville, Montreal, QC H3C 3J7, Canada.
Email: alexandra.lapierre@umontreal.ca
Results. From 962 unique records, 45 studies were included and 27 provided enough data on subjective cognitive load (i.e., Paas Scale and NASA-Task Load Index scores) to be meta-analyzed. In the multivariate analysis for the NASA-Task Load Index scores, each repetition of a simulation using the same scenario resulted in a linear decrease in cognitive load. In contrast, technology-based instruction before or during a simulation activity was associated with higher cognitive load. In the univariate analyses, other features such as feedback and instructor presence were also statistically associated with cognitive load. Regarding the univariate analyses of the Paas Scale scores, simulator type, briefing, debriefing, and repetitive practice were statistically associated with cognitive load.

Conclusion. This is the first meta-analysis exploring the relationship between clinical simulation design features and novice healthcare professionals’ cognitive load. Although the findings show that several design features can potentially increase or decrease cognitive load, several gaps and inconsistencies in the current literature make it difficult to appreciate how such reciprocity influences novice healthcare professionals’ learning. These limitations are discussed and avenues for educators and further research are suggested.

Keywords
Cognitive load, mental effort, simulation, healthcare education, meta-analysis

Background
Clinical simulations are often used in healthcare education to replicate real-world patient scenarios in an environment safe for education and experimentation purposes (Cook et al., 2013). According to simulation standards, a clinical simulation should be composed of three phases: (1) briefing, (2) simulation, and (3) debriefing (INACSL Standards Committee, Watts, et al., 2021). Specifically, the briefing is the time allocated before the simulation to prepare participants for the simulated clinical experience (INACSL Standards Committee, McDermott, et al., 2021). It can be used to identify participants’ learning expectations, build integrity, trust, and respect among them, or provide orientation to the simulation environment. The simulation phase is when participants experience the simulated clinical experience according to specific learning objectives (INACSL Standards Committee, Watts, et al., 2021). During this phase, a facilitator (i.e., generally a person skilled and trained in simulation) is dedicated to facilitating the simulated experience. They must adjust the course of the simulation according to learners and provide appropriate feedback during or after the simulation (INACSL Standards Committee, Watts, et al., 2021). The debriefing usually takes place after the simulation and is when professionals come together to reflect and discuss their experience for learning and improvement (INACSL Standards Committee, Decker, et al., 2021). Although widely used in healthcare professional education, clinical
Simulations remain complex interventions composed of several design features, that is, the active ingredients that should ensure the simulation effectiveness depending on the context (Cook et al., 2013). Indeed, each simulated clinical experience can also be performed according to different degrees of realism (high, medium, low fidelity), complexity (easy, moderate, hard), composition (individual, group), duration, and many more. However, the impact of these different design features on the cognitive learning process of healthcare professionals is still poorly understood.

Embedded in neurocognitive principles, recent neuroimaging studies have shown that clinical case scenarios with various complexity (easy/hard) elicited differential activation of regions of the human prefrontal cortex associated with working memory depending on the level of expertise of healthcare professionals, more so for healthcare professionals in training than experts (Hruska et al., 2016). This suggests that there could be an important relationship between clinical simulation design features and working memory, particularly for novice healthcare professionals (i.e., healthcare professionals in training or their early stage of clinical practice; less than five years). Thus, this idea can be further refined through the lens of the cognitive load (CL) theory (Paas & Van Merriënboer, 1994). CL theory is based on the understanding of human “cognitive architecture” or how we process information (Paas & Van Merriënboer, 1994). CL theory proposes that individuals have limited mental resources to deal with and integrate new information into long-term memory (Paas & Van Merriënboer, 1994). Specifically, there are two main types of CL: intrinsic load and extraneous load. In the context of clinical simulations, intrinsic load deals with elements of the simulations that are relevant to the learning task at hand, such as the learning objectives and theoretical content. Extraneous load is dedicated to the management of information that has the potential to distract healthcare professionals involved in the clinical simulations, such as noise or unclear directions (Leppink et al., 2014). Both loads are additive and may lead to cognitive overload or underload if the educational activity is not properly adapted to healthcare professionals’ level of expertise and information processing capacity.

The fundamental tenets of CL theory suggest that novice healthcare professionals cannot refer to as much knowledge and experience as experts when faced with the same clinical situation since they have not experienced the situation enough times to have stored the information in their long-term memory (Fraser et al., 2015; Szulewski et al., 2021). In addition, the novices do not have the experts’ ability to disregard redundant or irrelevant stimuli from the environment and, therefore, could be more easily distracted in a simulation activity that contains elements that are superfluous to the learning objectives, for example, an infusion pump that is constantly ringing (Szulewski et al., 2021). Some studies also suggest that clinical simulations can generate emotions, stress, and uncertainty among novice professionals, especially students, which would increase CL (mainly extraneous load) and interfere with their learning processes (Fraser & McLaughlin, 2019; Schlair et al., 2015; Szulewski et al., 2021). Accordingly, studies have shown that keeping CL low at the beginning of a simulation training curriculum and increasing it steadily by providing more support to the novice
healthcare professionals and gradually decreasing it—referring to the concept of scaffolding—can be beneficial for learning (Fraser et al., 2015; Josephsen, 2015). However, this is complex as the different simulation design features that are brought into interaction have the potential to influence more or less the novice healthcare professionals’ CL. In fact, we have little information on the association between simulation design features and novice healthcare professionals’ CL, which means that we know little about the activable levers to adjust the learning task and meet this subgroup’s cognitive specificities. We believe this exercise is essential for novice professionals as they are increasingly trained in their academic programs and early professional careers through clinical simulations (Aebersold, 2018; Datta et al., 2012).

To date, two literature reviews have synthesized the evidence on healthcare professionals’ CL in clinical simulations, and both identified several simulation design features that appear to increase or decrease CL (Rogers & Franklin, 2021; Sewell et al., 2019). However, results were presented narratively and included healthcare professionals from all levels of expertise, increasing the risk of bias in their interpretation. Moreover, the reviews combined the results of primary studies using subjective (e.g., self-reported rating scales) and objective (e.g., pupil fixation, heart rate) measures. Although these are all valid measures of CL, it can be argued that they are not equivalent conceptually. Indeed, subjective measures reflect CL as perceived by the learner at a given point in time (usually after the clinical simulation). In contrast, objective measures are based on observable behaviors or physiological parameters typically measured throughout the simulation (Brünken et al., 2010).

The present study aimed to perform the first meta-analysis of the relationship between clinical simulation design features and novice healthcare professionals’ CL while considering types of CL measures. This aim to better understand the complex and not entirely clear relationships between simulation design features and CL (Fraser et al., 2015). We believe that this is an essential first step to better understanding the mediating effect of CL on learning or performance in simulation. This could help identify what influence the novice healthcare professionals’ cognitive processing so that it can be further studied.

Methods

This systematic review was based on the Joanna Briggs Institute (JBI) methodology for systematic reviews of effectiveness (Tufanaru et al., 2017) and is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021). The review protocol was published (Lapierre et al., 2021) and registered prospectively in PROSPERO (CRD42020187723). Here, we describe changes made since the publication of the protocol and present an abridged version of the method.
Initially, we aimed to evaluate the effect of clinical simulations and their design features on CL in healthcare professionals and students. Due to a lack of studies comparing simulations with other interventions or comparing the design features of similar simulations, we could not answer this question. For this reason, we decided to focus on the associations between simulation design features and the CL of healthcare professionals. In addition, since we found that most of the studies in this area were conducted with novice healthcare professionals and based on our understanding of CL theory, we decided to concentrate on this specific population. The main research question that guided the present review and meta-analytic work was as follows: what are the associations between simulation design features and novice healthcare professionals’ CL?

Search strategy and data sources

The search strategy was based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses literature search extension (PRISMA-S) (Rethlefsen et al., 2021). We searched five databases in May 2020 for studies published in English or French in peer-reviewed journals: Cumulative Index to Nursing and Allied Health Literature (CINAHL; EBSCOhost), Excerpta Medica dataBASE (EMBASE; Ovid), Education Resources Information Center (ERIC; ProQuest), MEDLINE (Ovid), PsycINFO (APA PsycNET), and Web of Science—Science Citation Index (SCI) and Social Sciences Citation Index (SSCI; Institute for Scientific Information [ISI]—Thomson Scientific). Our search strategy combined thesaurus terms and text words related to 1) healthcare professionals and students; 2) simulation activities; and 3) CL (see Table S1). We also scanned the reference lists of eligible studies to identify additional relevant studies.

Eligibility criteria

Studies that met the criteria listed below were eligible for inclusion:

- Population: Novice healthcare professionals from any discipline (e.g., medicine, nursing, paramedic, pharmacy). More specifically, we included all novice professionals currently in training and not graduated, that is, students and medical residents, and also graduated professionals with less than five years of experience. Studies involving students/residents/novice professionals and more expert clinicians could be included only if results were stratified according to participants’ training status. In addition, novice healthcare professionals had to engage actively in simulation activities for their own learning (i.e., not as observers, actors, or educators).
- Intervention: Any non-digital simulation modality used to reproduce a patient encounter (e.g., procedural simulation with part-task trainers; simulated clinical
immersion with low- to high-fidelity manikins; simulated or standardized patients) (Chiniara et al., 2013).

- **Comparison:** Any other educational intervention or simulation activity, or no comparator.
- **Outcome:** Self-reported or objective CL measures during or after a simulation activity. Self-reported questionnaires were the primary outcome of this review, as they represent the most common method to measure CL in clinical simulation (Naismith & Cavalcanti, 2015). Objective indicators of CL were also considered, including 1) secondary task performance (e.g., reaction to a vibration stimulus); 2) physiological indices (e.g., heart, eye, or brain activity); and 3) observer ratings (e.g., changes in behaviors or speeches).
- **Design:** Experimental, quasi-experimental, or pre-experimental design (i.e., randomized controlled trials, non-randomized controlled trials, before and after studies, interrupted time-series studies, post-test only studies).
- **Context:** Any healthcare or educational setting.

**Exclusion criteria**

Digital simulations (e.g., virtual simulation or computer-based simulation) were excluded to reduce heterogeneity, as the CL of healthcare professionals in a digital environment may be affected by specific digital factors (e.g., computer literacy levels, digital device features).

**Study selection**

We worked independently and in duplicate to screen all titles and abstracts using Covidence (Veritas Health Innovation, Melbourne, Australia). We resolved disagreements by consensus or with the involvement of a third reviewer. We then assessed full texts for eligibility using the same method.

**Data extraction**

We developed a data extraction form based on the standardized JBI tool and tested it with three studies (see complete data extraction form in the published protocol) (Lapierre et al., 2021). Two reviewers independently extracted data related to study identification, methods, population, interventions (simulation activities and design features), comparators, and relevant outcomes. Then, one reviewer coded each clinical simulation activity from the included studies for the presence or absence of design features described in a framework adapted from Cook et al. (2013) (see Table 1). The unit of analysis was clinical simulation activity, meaning that studies involving several simulation activities were coded multiple times. A second reviewer validated coding. Participants’ characteristics were categorized as students, residents, or novice
professionals. We could not extract participants’ age since it was missing from most studies.

To gain feedback on our codification, we e-mailed all corresponding authors of included studies on two occasions to ask if it was representative of what occurred in their simulation activities. Furthermore, we asked them to provide missing outcome data, if applicable. In total, 14/45 authors (31%) responded.

Quality assessment

Two reviewers assessed study quality using standardized critical appraisal instruments from the JBI (Tufanaru et al., 2017). Based on a previous review (Roberts & Cooper, 2019), quality scores of 10/13 or above were considered high, 7 to 9 as moderate, and anything lower than 7 as low for randomized controlled trials; whereas quality scores of 8 or 9/9 were considered as high, 6 or 7 as moderate, and any scores lower than 6 as low for quasi-experimental studies.

Data analysis

In collaboration with a statistician, we used univariate and multivariate linear mixed models to describe the associations between simulation design features and novice healthcare professionals’ CL. To account for varying sample sizes, the results of studies were weighted using the methods of meta-analyses (i.e., the inverse of the variance). Moreover, since some participants were included in more than one simulation activity (i.e., more than one observation), the dependence of observations was modeled using a compound-symmetry covariance matrix. The dependent variables related to CL were treated as continuous. NASA-Task Load Index (NASA-TLX) scores were adjusted to a 100 scale when authors used a 0–120 scale. Most of the design features were treated as categorical (presence or absence), except for three features treated as continuous (i.e., repetitive practice, repeated exposure to simulations, and time allowed to the simulation activity). To build the multivariate linear mixed models, all independent variables that were significant in the univariate models were put in the equation. Those that were found nonsignificant in the multivariate models were then deleted one at a time until all remained statistically significant (Kellar & Kelvin, 2013).

All analyses were performed using IBM SPSS Statistics for Macintosh, Version 26.0 (IBM Corp., Armonk, NY). Results are expressed using unstandardized regression coefficients ($\beta$) with 95% confidence intervals. $P$ values $\leq$ 0.05 were considered significant. Analyses were performed when five or more simulation activities (i.e., observations) were available for the same design feature and outcome, using the same measurement tool. Although ten observations are generally recommended (Higgins & Thomas, 2019), we decided to reduce to five observations because results were often poorly reported, not allowing us to meta-analyze as many studies as expected. Finally, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths, we found no outliers in the data.
| Design features                  | Definition                                                                                                                                  | n (%) |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------|
| **Simulator type**              | **Device used during a simulation activity.**                                                                                             |       |
| High-fidelity manikin           | **Full body manikin that can imitate, at a very high level, the functions of the human body (e.g., breathing).**                         | 23 (51)|
| Part-task trainer               | **A model that is a representation of a part of the human body, such as an arm or an abdomen.**                                            | 19 (42)|
| Cadaver or live porcine         | **Dead animal body or live models (e.g., porcine).**                                                                                       | 4 (9) |
| Standardized patient            | **Person who has been coached to simulate an actual patient.**                                                                              | 4 (9) |
| Instructional design features   | **Person who has been coached to simulate an actual patient.**                                                                              |       |
| **Briefing**                    | **Activities conducted before the simulated scenario for participants to familiarize themselves with the simulation environment and objectives.** | 19 (42)|
| Scenario variation              | **Each participant is exposed to a series of different clinical scenarios.**                                                                 | 19 (42)|
| Repetitive practice             | **Repetition of a simulation activity with a similar scenario one or more times.**                                                          | 15 (33)|
| Repeated exposure               | **Participation in two or more sequential simulation activities with different scenarios or tasks.**                                         | 4 (9) |
| Range of task difficulty        | **Variation in the difficulty of the task throughout a sequence of simulation activities (explicitly stated).**                             | 15 (33)|
| Debriefing                      | **Interactions with a facilitator/instructor to review and reflect on the group’s or individual’s performance after the simulation scenario.** | 11 (24)|
| Feedback                        | **Information on performance received from an instructor, peer, or computer during or after the simulation activity (excluding debriefing).** | 11 (24)|
| Stressor                        | **Addition of stressors unrelated to the patient’s condition (e.g., interruptions, emotional stressors; needs to be explicitly stated).**   | 9 (20) |
| Roles                           | **Each participant is assigned with different roles in the simulation scenario (e.g., leader, team member)**                                | 6 (13) |
| Distributed practice            | **Training spread over time; involves >1 day of simulation training.**                                                                     | 6 (13) |
| Mastery learning                | **Training model in which clear standards of performance on a task must be attained before qualifying for the next task.**                 | 3 (7) |
| Individual composition          | **Person participating in the simulation**                                                                                                 |       |
| Individual                      | **One learner/participant actively involved in the simulation.**                                                                           | 34 (76)|
| Group                           | **Two or more learners/participants actively involved in the simulation.**                                                                    | 12 (27)|

(continued)
Table 1. (continued)

| Design features       | Definition                                                                 | n (%) |
|-----------------------|----------------------------------------------------------------------------|-------|
| Actor                 | A person trained to mimic the presence of other health professionals or family members throughout the simulation. | 14 (31) |
| Instructor            | A person who participates in the implementation and/or execution of the simulation without being actively involved in the scenario. | 22 (49) |
| Duration              | Time allowed for the scenario (excluding briefing and debriefing).         |       |
| 0–20 minutes          |                                                                            | 29 (64) |
| 21–40 minutes         |                                                                            | 5 (11)  |
| 41–60 minutes         |                                                                            | 4 (9)   |
| Added modality        | Addition of one or more instructional modalities to a simulation activity. |       |
| Technology-based      | Educational strategy to present steps related to the task or actions (e.g., practice guidelines) using an electronic format (e.g., video). | 11 (24) |
| instruction           |                                                                            |       |
| Instruction           | Educational strategy to present steps related to the task or actions (e.g., practice guidelines) not using an electronic format (e.g., paper-based algorithm). | 8 (18)  |
| Lecture               | Formal didactic session delivered before an audience or class, often more focused on theory. | 3 (7)   |
| Team training         | Instruction on teamwork skills (e.g., TeamSTEPPS; Canadian Patient Safety Institute, 2018) | 1 (2)   |
| Sensory augmentation  | Addition of a feature or effect to enhance the sensory experience.          |       |
| Auditory              | Addition of an auditory element (e.g., alarm).                             | 6 (13)  |
| Visual                | Addition of a visual element (e.g., light, image).                         | 3 (7)   |

Adapted from Cook et al. (2013) and the Healthcare Simulation Dictionary (Lopreiato, 2016).

Results

The initial search yielded 962 unique records (Figure 1). After screening the titles and abstracts, the full text of 180 potentially eligible studies was screened, and 45 studies were included in this review.

Characteristics of the studies and participants

Most studies took place in North America (n = 29, 64%) between 2011 and 2021 (n = 40, 89%) with participants from medicine (n = 36, 80%), nursing (n = 5, 11%), pharmacy (n = 2, 4%), and pre-hospital services (n = 1, 2%). Only 1 study (2%) included participants from multiple disciplines. Most studies were conducted with
students (n = 28, 62%) or residents (n = 19, 42%); only 1 study (2%) was conducted with novice graduated professionals (i.e., less than 5 years of clinical experience). The median sample size of the included studies was 30 (range 8–299).

Most studies were randomized controlled trials with parallel groups (n = 14, 31%) or crossover designs (n = 9, 20%). Thirteen studies (29%) were single-group repeated measures, 5 (11%) were single-group post-test only, and 4 studies (9%) were non-randomized controlled trials. Most studies (n = 37, 82%) used self-reported questionnaires such as the NASA-TLX (n = 20, 45%; Hart & Staveland, 1988), the Paas

**Figure 1.** Flow diagram of study selection.
Scale (n = 13, 29%; Paas, 1992), the Surgery-Task Load Index [SURG-TLX] (n = 2, 4%; Wilson et al., 2011), the Cognitive Load Questionnaire (n = 3, 7%; Leppink et al., 2013), or a questionnaire specifically developed for the study (n = 1, 2%). Secondary task analysis (n = 10, 22%) were also used, including reaction time (n = 6, 13%), detection rate (n = 2, 4%), number of chart errors (n = 2, 4%), and math calculation (n = 1, 2%). Physiologic measures (n = 7, 16%) such as eye tracking (n = 6, 13%), heart rate (n = 4, 9%), electrodermal activity (n = 1, 2%) were less frequent. Only 1 study (2%) used behaviors as a measure of CL. Nine studies (20%) used multiple CL measures.

Characteristics of the studies can be found in Table S2.

Methodological quality assessment

Regarding randomized controlled trials (n = 22), 6 (26%), 10 (43%), and 6 (26%) were rated as high, moderate, and low quality, respectively. Regarding quasi-experimental studies (n = 23), 6 (26%), 12 (52%), and 5 (22%) were of high, moderate, and low quality, respectively (see Table S3-S4).

Association between simulation design features and CL

Table 1 presents the definition of the simulation design features and their prevalence in the 45 studies according to our codification. The most common simulation design features were the simulator type, specifically high-fidelity manikin (n = 23, 51%) and part-task trainer (n = 19, 42%), the presence of briefing (n = 19, 42%), the variation in clinical scenarios (n = 19, 42%), and the simulation composition, which mainly was individual (n = 34, 76%).

Twenty-seven studies provided enough data on the NASA-TLX (n = 15; 49 unique observations) and the Paas Scale (n = 13; 46 unique observations) to be meta-analyzed; one study used both instruments. The other studies could not be meta-analyzed due to a lack of data from the same measurement methods.

NASA-TLX. In the 49 observations using the NASA-TLX, the mean score was 53.19 with a standard deviation of 16.39 (range 0–100). As shown in Figures 2 and 3a, the univariate analyses revealed significant associations (P ≤ .05) between the NASA-TLX scores and four design features: 1) instructor or simulator feedback; 2) presence of an instructor; 3) technology-based instructions before or during the simulation activity; and 4) repetitive practice (i.e., opportunity to repeat a simulation with a similar scenario). Specifically, feedback during or after the simulation and the presence of an instructor were associated with lower CL (β = -18.33, 95% CI -28.28, -8.38, P = 0.001; β = -12.11, 95% CI -21.78, -2.45, P = 0.015). The presence of technology-based instruction was associated with higher CL (β = 16.18, 95% CI -7.86, 24.50, P = 0.001), whereas CL decreased linearly with each repetition of a simulation with the same scenario or task (β = -10.88, 95% CI -18.65, -3.11, P = .020; Figure 3a). Downward trends in CL were also observed with repeated exposure to simulation (i.e., two or more sequential simulation activities with
different scenarios or tasks) and increasing time allowed (i.e., in 10-min increments), but they did not reach statistical significance ($\beta = -8.01$, 95% CI -18.14, 2.11, $P = 0.115$; $\beta = -10.37$, 95% CI -25.79, 5.06, $P = 0.157$). In the mixed model analysis, the presence of technology-based instruction ($\beta = 7.13$, 95% CI, 3.69, 10.56, $P < .001$) and repetitive practice ($\beta = -7.35$, 95% CI -8.58, -5.11, $P < .001$) remained significant.

**Paas Scale.** In the 46 observations using the Paas Scale, the mean score was 6.17 with a standard deviation of 1.43 (range 1–9). As shown in Figures 3b and 4, results from the univariate analysis for 46 unique observations revealed significant associations ($P \leq 0.05$) between the Paas Scale scores and four design features: 1) simulator type (i.e., high-fidelity manikin compared to part-task trainer); 2) presence of a briefing; 3) presence of a debriefing; and 4) repetitive practice. Specifically, the use of high-fidelity manikin compared to a part-task trainer and the presence of a debriefing were associated with higher CL ($\beta = 1.31$, 95% CI 0.03, 2.59, $P = 0.046$; $\beta = 1.45$, 95% CI 0.25, 2.65, $P = 0.021$). However, the presence of a briefing was associated with lower CL ($\beta = -1.27$, 95% CI -2.48, -0.06, $P = 0.040$). For repetitive practice, the results showed that with each opportunity to repeat a simulation with a similar scenario or task, CL decreased linearly ($\beta = -0.22$, 95% CI -0.45, -0.00, $P = 0.047$; Figure 3b). The presence of technology-based instruction was very close to being statistically significant and showed an increasing trend ($\beta = 1.26$, 95% CI -0.03, 2.55, $P = 0.054$). Repeated exposure could not be analyzed due to a lack of data. In the mixed model analysis, no variable remained significant, potentially due to a lack of power.

**Discussion**

In this review, we explored the association between clinical simulation design features and subjective measures of CL. Simulation studies using objective measures

![Figure 2. Association between simulation design features and NASA-TLX scores.](image-url)
of CL were insufficient in number for meta-analysis. Although the state of literature limits current results, there are essential points to discuss regarding design features and the measurement of CL in clinical simulation studies. The following discussion focuses on research gaps and recommendations for the field of clinical simulation design and beyond.

**Figure 3.** Association between repetitive practice and NASA-TLX/Paas Scale scores
Simulation design features

Repetitive practice was the simulation design feature that appears to be most strongly related to CL when measured by different self-reported questionnaires. Specifically, many studies have shown that repeating a simulation scenario decreases the perceived CL of novice healthcare professionals (Abe et al., 2019; Haji, Khan, et al., 2015; Haji, Rojas, et al., 2015; Hu et al., 2016; Jirapinyo et al., 2017; Yurko et al., 2010). This could be explained by novices’ increased familiarity with the scenario or perception that their performance improves through repetition (Haji, Khan, et al., 2015; Young et al., 2014). This could also be explained by the fact that several task repetitions can lead to automaticity (i.e., the ability to execute with less effort and fewer mental resources; Paas & Van Merriënboer, 1994; Stefanidis et al., 2008). However, in this review, repeated exposure to simulation, regardless of the scenario or task also tended to diminish CL even though this relationship did not reach statistical significance. This suggests that professionals were getting accustomed to the concept and functioning of simulation as an educational technique, which may also explain why their CL eventually declined over time.

The presence of technology-based instruction—mainly video demonstrations before the clinical simulation—strongly increased novices’ self-reported CL on the NASA-Tlx, while it was very close to being significant for the Paas Scale. In recent years, educational technology has become more accessible and popular in healthcare professions education (Car et al., 2019; Grimwood & Snell, 2020; Kononowicz et al., 2019). Yet, we are not aware of any study that specifically examined the impact of educational technology before or during a clinical simulation on CL. However, there is an extensive literature on CL and technological-based education, such as multimedia learning environments (Kalyuga & Liu, 2015; Mutlu-Bayraktar et al., 2019; Sweller, 2020), or even digital simulations (Chang et al., 2021; Frederiksen et al., 2020)—which
was excluded from this review because of the significant difference between the training modalities. Still, an important message is that it is not the technology itself that matters, but how and when it is used to optimally engage the learners’ cognition (Kalyuga & Liu, 2015; Sweller, 2020). Our results highlight the importance of reflecting on the educational principles and value of such technological additions, which could be the object of another research question on its own.

The remaining design features were only significant in the univariate analyses. Feedback was associated with lower NASA-TLX scores. Feedback is related to the information on performance received from an instructor, peer, or computer during or after the simulation activity (Cook et al., 2013). Several studies included in this review have specifically focused on different aspects of instructor feedback, such as the use of video, positive vs. negative feedback, or its frequency (Abbott et al., 2017; Aljamal et al., 2019; Bosse et al., 2015; Byrne et al., 2002; Tohil et al., 2020). In these studies, CL seemed to be similar or decrease across different feedback groups, which is consistent with our results. Indeed, the feedback given by an instructor would allow to better structure the simulated task so that it is better understood (Hatala et al., 2014), thus decreasing the intrinsic CL. In contrast, two studies evaluated the effects of simulator feedback (Brown et al., 2018; Hempel et al., 2019), while no studies evaluated feedback from peers. Interestingly, the specific results of the simulator feedback showed the opposite effect from instructor feedback on CL. As Hatala et al. (2014), more theoretically based studies that include CL measures are still needed to better understand the benefits of this important design feature.

Instructor presence was also associated with lower NASA-TLX scores. In our opinion, these results are strongly associated with the presence of feedback, as an instructor is absolutely required to provide feedback. Indeed, no study has specifically looked at the presence vs. the absence of an instructor or even at instructors’ characteristics or behaviors that might influence learners’ CL. Indeed, as Frisby et al. (2018) did in a classroom context, it might be interesting to look at instructors’ behaviors that might influence the extraneous CL of novice healthcare professionals. This study showed that behaviors such as speaking quickly or using offensive words could impact learners’ extraneous CL.

Regarding the Paas Scale, simulator type (i.e., high-fidelity manikin compared to part-task trainer) was associated with higher scores. This is consistent with the postulate that simulation fidelity is positively correlated to intrinsic CL (Reedy, 2015). In fact, it is increasingly recognized that to manage novices’ CL better, you need to start with lower fidelity and increase it over time (Brydges et al., 2010; Reedy, 2015). Indeed, the idea that an immersive, high-fidelity learning environment is ideal for preparing novice healthcare professionals for clinical practice is increasingly questioned (Reedy, 2015). However, since fidelity is a complex concept composed of several dimensions (i.e., physical, psychological, conceptual; Paige & Morin, 2013), these results do not allow us to differentiate the impact of the different spheres of fidelity on CL.

The presence of a briefing was associated with lower Paas Scale scores. Best practices emphasize the importance of briefing participants before clinical simulations
since it allows professionals to familiarize themselves with the simulation environment and better understand the activity’s objectives (INACSL Standards Committee, Decker, et al., 2021). Consequently, it could have an influence on the extraneous CL, as this practice allows presenting and clarifying the different elements that could potentially distract the learner from the learning task. In fact, these results align with the numerous studies that showed positive outcomes of briefing for healthcare professionals (Tyerman et al., 2019).

As opposed to a briefing, the presence of a debriefing was associated with higher Paas Scale scores. The only study that clearly aimed to examine the effect of debriefing on CL showed that medical students had higher CL scores after the debriefing than after the scenario (Fraser & McLaughlin, 2019). However, another study, using an objective measure, showed that surgical residents who were debriefed had a reduced median response time to a vibration stimulus compared to those who were not—suggesting reduced CL as a result of debriefing (Boet et al., 2017). This apparent contradiction could be explained by the difference in CL measurement (subjective vs. objective), the difference in expertise (medical students vs. residents), or the use of different debriefing methods (Promoting Excellence And Reflective Learning in Simulation [PEARLS] (Eppich & Cheng, 2015) vs. Debriefing with Good Judgment (Rudolph et al., 2007)), among other factors.

Although several associations did not reach statistical significance, some trends are worth noting. First, complex simulations led to higher CL compared to simpler simulations, which surprisingly did not reach statistical significance even if all studies followed the same trend (Haji, Khan, et al., 2015; Kataoka et al., 2011; Tremblay et al., 2019). Indeed, the idea that novice healthcare professionals should be confronted from the beginning with the full complexity of clinical situations to be well prepared for the clinical reality, is now discouraged according to CL theory (Reedy, 2015). Second, sensory augmentation at the auditory and visual level also seemed to increase CL on the Paas Scale, which is consistent with findings that auditory interruptions (e.g., telephone call, monitor alarms) increase the CL of novice healthcare professionals (Campoe & Giuliano, 2017; Weigl et al., 2016; Wucherer et al., 2015). Other studies have investigated the effect of several concurrent stimuli, making it difficult to isolate their influence on CL (Mills et al., 2016; Tremblay et al., 2017).

**Measuring CL**

Subjective CL measures were the most frequent in the included studies, possibly because they are easy to use and non-intrusive. Although the Paas Scale and NASA-TLX have been used interchangeably, they somewhat differ in their form and conceptual foundations (Naismith et al., 2018). The Paas Scale is a single-item instrument that estimates mental effort on a 9-point scale, whereas the NASA-TLX consists of six subdomains scored on a 0–20 scale: mental demand, physical demand, temporal demand, effort, frustration, and performance (Hart & Staveland, 1988; Paas, 1992). A review by Naismith et al. (2015) found poor agreement between the two instruments in
a simulation setting—the Paas Scale correlated only with the mental demand sub-domain of the NASA-TLX. It is therefore not surprising that the results of the current meta-analyses were different for the two instruments.

Furthermore, an important limitation of these instruments is that they only capture the overall CL and not the contribution of each CL type (Brünken et al., 2010; Leppink et al., 2014). This is problematic since high or low overall CL is a very imprecise measure that does not allow to know which type of CL (intrinsic or extraneous) is impacted. To date, most instruments have been constructed to measure overall CL, and only a few measure intrinsic and extraneous loads separately (Klepsch et al., 2017; Leppink et al., 2013)—in this review, only four studies measured the different CL types (Josephsen, 2018; Naismith et al., 2015; Tremblay et al., 2017; Tremblay et al., 2019).

The complexity in developing this kind of instrument might be explained by the ongoing debate about what composes CL, especially questions regarding the “germane” load (Leppink & van den Heuvel, 2015). The germane load was identified in early versions of the theory as the construction and automation of mental schema and its integration into the long-term memory (Paas et al., 2003). However, due to incongruence on the theoretical level, the germane load was recently reconceptualized as a subtype of the intrinsic load (Jiang & Kalyuga, 2020; Kalyuga, 2011; Leppink & van den Heuvel, 2015). For this reason, authors who aimed to measure different CL types had few options. In fact, they had to develop or adapt the rare instruments measuring the CL types, such as the Cognitive Load Questionnaire of Leppink et al. (2013), because they are not adapted to the simulation context. Therefore, it is understandable that many researchers have turned to instruments already validated and easy to use, such as the Paas Scale and the NASA-TLX.

Also worth noting, few studies have used objective measures of CL. An advantage of objective measures is that they allow measuring CL during the clinical simulation and observing temporal variations (Zheng & Greenberg, 2018). However, these measures have likely been used less because they may compromise the authenticity of a simulation or may interfere physically with the participants (e.g., having to react to a vibration stimulus during the simulation, adding an eye-tracking device; Zheng & Greenberg, 2018). Unobtrusive methods, such as behavioral observations, are beginning to be developed and show a promising avenue—however, research is still limited (Naismith et al., 2018).

**Recommendations for educators and researchers**

The results suggest that repeated exposure to simulations could be a successful strategy for educators who aim to reduce novices’ CL. Otherwise, we recommend that educators carefully reflect on the necessity of some design features before including them in simulations, particularly for technologies (e.g., high-fidelity mannequins or videos). Are these technologies absolutely essential to achieve the learning objectives? If so, at what time the technologies should be optimally introduced during the simulated clinical experience? In addition, until proven otherwise, instructional strategies known to be
central to simulations, such as briefing, feedback, and debriefing, should be constantly used, as per the simulation best practices (INACSL Standards Committee, Watts, et al., 2021).

For researchers, we believe that one of the needs to advance knowledge in this field is to develop, adapt, and use an instrument suitable for assessing the different types of CL in clinical simulation context. For example, Tremblay et al. (2019) have begun to work on adapting the Cognitive Load Questionnaire of Leppink et al. (2013), but its validation and wider dissemination remain to be done. Still in terms of measures, other non-subjective and least intrusive methods should be developed (such as behavioral observations), to address the specific context of clinical simulations, and to allow for the use of diverse data collection methods—ensuring stronger internal validity of studies. Also, we believe that rigorous control group studies evaluating the major simulation design features such as briefing, debriefing and feedback on CL are needed to understand the association between these variables better. To this end, it is essential to properly describe the simulation interventions being evaluated—Reporting guidelines for health care simulation research (Cheng et al., 2016) should be used to ensure consistency in the description of these complex educational interventions.

Strengths and limitations

This review has several strengths, including the implementation of a rigorous process based on the JBI and the new PRISMA guidelines. The protocol was prospectively registered, which enhances transparency. Although we had initially planned to evaluate the impact of simulation design features on CL through studies in which different clinical simulations were compared head-to-head, the absence of such studies precluded us from doing so. Thus, we had to rely on findings from univariate linear mixed models, which may be subject to confounding biases. Another limitation to this approach is that the multiplicity of statistical testing may have increased the risk of type 1 error. Despite these limitations, we firmly believe that the results, even if exploratory in nature, advance knowledge and put forward several research avenues on a subject of utmost importance in novice healthcare professions’ education. Although two independent reviewers performed coding with good agreement, it was based on the description of the clinical simulations, which could be subject to reporting bias. Finally, the results are based on studies that are generally of moderate methodological quality. For all these reasons, these results should be considered with caution.

Conclusion

This review provides the first meta-analysis exploring the relationships between the simulation design features and the novice healthcare professionals’ CL. Overall, the results demonstrate that there are significant associations between several simulation design features and novices’ CL, including instructor or simulator feedback, presence of an instructor, technology-based instructions, repetitive practice, simulator type, and
debriefing. However, the results also highlight that there are still many unanswered questions to understand better these relationships, such as how to measure the different CL types appropriately. The lack of instruments leads to fundamental difficulties in the generation of new knowledge in this field. Ultimately, this work aims to understand better the relationships between the simulation design features and CL and then examine the impact these relationships have on the learning of novice healthcare professionals—but before we get to that point, much work remains to be done.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

Alexandra Lapierre  ©️ [https://orcid.org/0000-0002-8704-4940](https://orcid.org/0000-0002-8704-4940)

**Supplemental Material**

Supplemental material for this article is available online

**References**

Abbott, E. F., Thompson, W., Pandian, T. K., Zendejas, B., Farley, D. R., & Cook, D. A. (2017). Personalized Video Feedback and Repeated Task Practice Improve Laparoscopic Knot-Tying Skills: Two Controlled Trials. *Academic Medicine, 92*(11), S26-S32. [https://doi.org/10.1097/acm.0000000000001924](https://doi.org/10.1097/acm.0000000000001924)

Abe, T., Dar, F., Amnattrakul, P., Aydin, A., Raison, N., Shinohara, N., Khan, M. S., Ahmed, K., & Dasgupta, P. (2019). The effect of repeated full immersion simulation training in ureterorenoscopy on mental workload of novice operators. *BMC Medical Education, 19*(1), 318. [https://doi.org/10.1186/s12909-019-1752-2](https://doi.org/10.1186/s12909-019-1752-2)

Aebersold, M. (2018). Simulation-Based Learning: No Longer a Novelty in Undergraduate Education”. *The Online Journal of Issues in Nursing, 23*(2). [https://doi.org/10.3912/OJIN.Vol23No02PPT39](https://doi.org/10.3912/OJIN.Vol23No02PPT39)

Aljamal, Y., Prabhakar, N., Saleem, H., & Farley, D. R. (2019). Can the Perceived Difficulty of a Task Enhance Trainee Performance? *Journal of Surgical Education, 76*(6), e193-e198. [https://doi.org/10.1016/j.jsurg.2019.08.005](https://doi.org/10.1016/j.jsurg.2019.08.005)

Boet, S., Sharma, B., Pigford, A. A., Hladkowicz, E., Rittenhouse, N., & Grantcharov, T. (2017). Debriefing decreases mental workload in surgical crisis: A randomized controlled trial. *Surgery, 161*(5), 1215-1220. [https://doi.org/10.1016/j.surg.2016.11.031](https://doi.org/10.1016/j.surg.2016.11.031)
Bosse, H. M., Mohr, J., Buss, B., Krautter, M., Weyrich, P., Herzog, W., Junger, J., & Nikendei, C. (2015). The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. *BMC Medical Education, 15*, 22. https://doi.org/10.1186/s12909-015-0286-5

Brown, L. L., Lin, Y. Q., Tofil, N. M., Overly, F., Duff, J. P., Bhanji, F., Nadkarni, V. M., Hunt, E. A., Bragg, A., Kessler, D., Bank, I., & Cheng, A., Int Network Simulation-based. (2018). Impact of a CPR feedback device on healthcare provider workload during simulated cardiac arrest. *Resuscitation, 130*, 111-117. https://doi.org/10.1016/j.resuscitation.2018.06.035

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring Cognitive Load. In J. L. Plass, R. Brünken, & R. Moreno (Eds.), *Cognitive Load Theory* (pp. 181-202). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.011

Brydges, R., Carnahan, H., Rose, D., Rose, L., & Dubrowski, A. (2010). Coordinating Progressive Levels of Simulation Fidelity to Maximize Educational Benefit. *Academic Medicine, 85*(5). https://doi.org/10.1097/ACM.0b013e3181d7aadbd

Byrne, A. J., Sellen, A. J., Jones, J. G., Aitkenhead, A. R., Hussain, S., Gilder, F., Smith, H. L., & Ribes, P. (2002). Effect of videotape feedback on anaesthetists’ performance while managing simulated anaesthetic crises: a multicentre study. *Anaesthesia, 57*(2), 176-179. https://doi.org/10.1046/j.1365-2044.2002.02361.x

Canadian Patient Safety Institute. (2018). *Step up your teamwork and communication with TeamSTEPPS Canada™*. https://www.patientsafetyinstitute.ca/fr/NewsAlerts/News/Pages/Step-up-your-teamwork-and-communication-with-TeamSTEPPS-Canada-2018-03-13.aspx

Campoe, K. R., & Giuliano, K. K. (2017). Impact of Frequent Interruption on Nurses’ Patient-Controlled Analgesia Programming Performance. *Human Factors, 59*(8), 1204-1213. https://doi.org/10.1177/0018720817732605

Car, J., Carlstedt-Duke, J., Tudor Car, L., Posadzki, P., Whiting, P., Zary, N., Atun, R., Majeed, A., & Campbell, J. (2019). Digital Education in Health Professions: The Need for Overarching Evidence Synthesis [Viewpoint]. *Journal of Medical Internet Research, 21*(2), e12913. https://doi.org/10.2196/12913

Chang, H.-Y., Wu, H.-F., Chang, Y.-C., Tseng, Y.-S., & Wang, Y.-C. (2021). The effects of a virtual simulation-based, mobile technology application on nursing students’ learning achievement and cognitive load: Randomised controlled trial. *International Journal of Nursing Studies*, 103948. https://doi.org/10.1016/j.ijnurstu.2021.103948

Cheng, A., Kessler, D., Mackinnon, R., Chang, T. P., Nadkarni, V. M., Hunt, E. A., Duval-Arnould, J., Lin, Y., Cook, D. A., Pusic, M., Hui, J., Moher, D., Egger, M., & Auerbach, M., for the International Network for Simulation-based Pediatric Innovation, R., & Education Reporting Guidelines, I. (2016). Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Advances in Simulation, 1*(1), 25. https://doi.org/10.1186/s41077-016-0025-y

Chiniara, G., Cole, G., Brisbin, K., Huffman, D., Cragg, B., Lamacchia, M., & Norman, D. (2013). Simulation in healthcare: a taxonomy and a conceptual framework for instructional design and media selection. *Medical Teacher, 35*(8), e1380-1395. https://doi.org/10.3109/0142159x.2012.733451
Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. *Medical Teacher, 35*(1), e867-898. https://doi.org/10.3109/0142159x.2012.714886

Datta, R., Upadhyay, K., & Jaideep, C. (2012). Simulation and its role in medical education. *Med J Armed Forces India, 68*(2), 167-172. https://doi.org/10.1016/s0377-1237(12)60040-9

Eppich, W., & Cheng, A. (2015). Promoting Excellence and Reflective Learning in Simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. *Simul Healthc, 10*(2), 106-115. https://doi.org/10.1097/0000000000000072

Fraser, K., & McLaughlin, K. (2019). Temporal pattern of emotions and cognitive load during simulation training and debriefing. *Medical Teacher, 41*(2), 184-189. https://doi.org/10.1080/0142159X.2018.1459531

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive Load Theory for the design of medical simulations. *Simulation in Healthcare, 10*(5), 295-307. https://doi.org/10.1097/SIH.0000000000000097

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., & Andersen, S. A. W. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical Endoscopy, 34*(3), 1244-1252. https://doi.org/10.1007/s00464-019-06887-8

Frisby, B. N., Sexton, B. T., Buckner, M. M., Beck, A.-C., & Kaufmann, R. (2018). Peers and Instructors as Sources of Distraction from a Cognitive Load Perspective. *International Journal for the Scholarship of Teaching and Learning, 12*(2), 6. https://doi.org/10.20429/ijslot.2018.120206

Grimwood, T., & Snell, L. (2020). The use of technology in healthcare education: a literature review. *MedEdPublish, 9*(1), 137. https://doi.org/10.15694/mep.2020.000137.1

Haji, F. A., Khan, R., Regehr, G., Drake, J., Ribauipierre, S., Dubrowski, A., Haji, F. A., & de Ribaupierre, S. (2015). Measuring cognitive load during simulation-based psychomotor skills training: sensitivity of secondary-task performance and subjective ratings. *Advances in Health Sciences Education, 20*(5), 1237-1253. https://doi.org/10.1007/s10459-015-9599-8

Haji, F. A., Rojas, D., Childs, R., de Ribaupierre, S., & Dubrowski, A. (2015). Measuring cognitive load: performance, mental effort and simulation task complexity. *Medical Education, 49*(8), 815-827. https://doi.org/10.1111/medu.12773

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. North Holland Press. https://doi.org/10.1016/S0166-4115(08)62386-9

Hatala, R., Cook, D. A., Zendejas, B., Hamstra, S. J., & Brydges, R. (2014). Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. *Advances in Health Sciences Education, 19*(2), 251-272. https://doi.org/10.1007/s10459-013-9462-8

Hempel, G., Heinke, W., Struck, M. F., Piegeler, T., & Rotzoll, D. (2019). Impact of Quantitative Feedback via High-Fidelity Airway Management Training on Success Rate in Endotracheal
Lapierre et al. 559

Intubation in Undergraduate Medical Students-A Prospective Single-Center Study. *Journal of Clinical Medicine*, 8(9), 14. https://doi.org/10.3390/jcm8091465

Higgins, J. P. T., & Thomas, J. (2019). *Cochrane Handbook for Systematic Reviews of Interventions* (2 ed.). Wiley Blackwell.

Hruska, P., Krigolson, O., Coderre, S., McLaughlin, K., Cortese, F., Doig, C., Beran, T., Wright, B., & Hecker, K. G. (2016). Working memory, reasoning, and expertise in medicine—insights into their relationship using functional neuroimaging. *Advances in Health Sciences Education, 21*(5), 935-952. https://doi.org/10.1007/s10459-015-9649-2

Hu, J., Lu, J., Tan, W., Lomanto, D., Hu, J. S. L., & Tan, W. B. (2016). Training improves laparoscopic tasks performance and decreases operator workload. *Surgical Endoscopy, 30*(5), 1742-1746. https://doi.org/10.1007/s00464-015-4410-8

Decker, S., Alinier, G., Crawford, S. B., Gordon, R. M., Jenkins, D., & Wilson, C., INACSL Standards Committee. (2021). Healthcare Simulation Standards of Best PracticeTM The Debriefing Process. *Clinical Simulation in Nursing, 58*, 27-32. https://doi.org/10.1016/j.ecns.2021.08.011.

McDermott, D. S., Ludlow, J., Horsley, E., & Meakim, C., INACSL Standards Committee. (2021). Healthcare Simulation Standards of Best PracticeTM Prebriefing: Preparation and Briefing. *Clinical Simulation in Nursing, 58*, 9-13. https://doi.org/10.1016/j.ecns.2021.08.008.

Watts, P. I., McDermott, D. S., Alinier, G., Charnetski, M., Ludlow, J., Horsley, E., Meakim, C., & Nawathe, P. A., INACSL Standards Committee. (2021). Healthcare Simulation Standards of Best PracticeTM Simulation Design. *Clinical Simulation in Nursing, 58*, 14-21. https://doi.org/10.1016/j.ecns.2021.08.009.

Jiang, D., & Kalyuga, S. (2020). Confirmatory Factor Analysis of Cognitive Load Ratings Supports a Two-Factor Model. *The Quantitative Methods for Psychology, 16*(3), 216-225. https://doi.org/10.20982/tqmp.16.3.p216

Jirapinyo, P., Abidi, W. M., Aihara, H., Zaki, T., Tsay, C., Imaeda, A. B., & Thompson, C. C. (2017). Preclinical endoscopic training using a part-task simulator: learning curve assessment and determination of threshold score for advancement to clinical endoscopy. *Surgical Endoscopy, 31*(10), 4010-4015. https://doi.org/10.1007/s00464-017-5436-x

Josephsen, J. (2015). Cognitive Load Theory and Nursing Simulation: An Integrative Review. *Clinical Simulation in Nursing, 11*(5), 259-267. https://doi.org/10.1016/j.ecns.2015.02.004

Josephsen, J. (2018). Cognitive Load Measurement, Worked-Out Modeling, and Simulation. *Clinical Simulation in Nursing, 23*, 10-15. https://doi.org/10.1016/j.ecns.2018.07.004

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review, 23*(1), 1-19. https://doi.org/10.1007/s10648-010-9150-7

Kalyuga, S., & Liu, T.-C. (2015). Guest editorial: Managing cognitive load in technology-based learning environments. *Journal of Educational Technology & Society, 18*(4), 1-8.

Kataoka, J., Sasaki, M., & Kanda, K. (2011). Effects of mental workload on nurses’ visual behaviors during infusion pump operation. *Japan Journal of Nursing Science, 8*(1), 47-56. https://doi.org/10.1111/j.1742-7924.2010.00158.x

Kellar, S. P., & Kelvin, E. A. (2013). *Munro’s Statistical methods for healthcare research* (6 ed.). Lippincott William & Wilkins
Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load [Original Research]. Frontiers in Psychology, 8(1997). https://doi.org/10.3389/fpsyg.2017.01997

Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakarou, N., Davies, D., Saxena, N., Tudor Car, L., Carlstedt-Duke, J., Car, J., & Zary, N. (2019). Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. Journal of Medical Internet Research, 21(7), e14676. https://doi.org/10.2196/14676

Lapierre, A., Arbour, C., Maheu-Cadotte, M. A., Radermaker, M., Fontaine, G., & Lavoie, P. (2021). Effect of simulation on cognitive load in health care professionals and students: protocol for a systematic review and meta-analysis. JBI Evidence Synthesis. https://doi.org/10.11124/jbies-20-00213

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. Behavior Research Methods, 45(4), 1058-1072. https://doi.org/10.3758/s13428-013-0343-1

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. Learning and Instruction, 30, 32-42. https://doi.org/10.1016/j.learninstruc.2013.12.001

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. Perspectives on Medical Education, 4(3), 119-127. https://doi.org/10.1007/s40037-015-0192-x

Lopreiato, J. O. (2016). Healthcare Simulation Dictionary (Vol. 16(17)-0043). Agency for Healthcare Research and Quality.

Mills, B. W., Carter, O. B., Rudd, C. J., Claxton, L. A., Ross, N. P., & Strobel, N. A. (2016). Effects of low- versus high-fidelity simulations on the cognitive burden and performance of entry-level paramedicine students: A mixed-methods comparison trial using eye-tracking, continuous heart rate, difficulty rating scales, video observation and interviews. Simulation in Healthcare, 11(1), 10-18. https://doi.org/10.1097/SIH.0000000000000119

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. Computers & Education, 141, 103618. https://doi.org/10.1016/j.compedu.2019.103618

Naismith, L. M., & Cavalcanti, R. B. (2015). Validity of cognitive load measures in simulation-based training: A systematic review. Academic Medicine, 90(11 Suppl), S24-35. https://doi.org/10.1097/ACM.0000000000000893

Naismith, L. M., Cavalcanti, R. B., & Haji, F. A. (2018). Measuring cognitive load in medical and health professions simulation training. In R. Zheng (Ed.), Cognitive Load Measurement and Application. Routledge. https://doi.org/10.4324/9781315296258

Naismith, L. M., Cheung, J. J., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. Medical Education, 49(8), 805-814. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med12&AN=26152492
Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist, 38*(1), 63-70. https://doi.org/10.1207/S15326985EP3801_8

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371. https://doi.org/10.1007/BF02213420

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Paige, J., & Morin, K. (2013). Simulation Fidelity and Cueing: A Systematic Review of the Literature. *Clinical Simulation in Nursing, 9*, e481–e489. https://doi.org/10.1016/j.ecns.2013.01.001

Reedy, G. B. (2015). Using Cognitive Load Theory to Inform Simulation Design and Practice. *Clinical Simulation in Nursing, 11*(8), 355-360. https://doi.org/10.1016/j.ecns.2015.05.004

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., Blunt, H., Brigham, T., Chang, S., Clark, J., Conway, A., Couban, R., de Kock, S., Farrah, K., Fehrmann, P., Foster, M., Fowler, S. A., Glanville, J., & Group, P.-S. (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews, 10*(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Roberts, F., & Cooper, K. (2019). Effectiveness of high fidelity simulation versus low fidelity simulation on practical/clinical skill development in pre-registration physiotherapy students: a systematic review. *JBI Evidence Synthesis, 17*(6), 1229-1255. https://doi.org/10.11124/jbisrir-2017-003931

Rogers, B. A., & Franklin, A. E. (2021). Cognitive load experienced by nurses in simulation-based learning experiences: An integrative review. *Nurse Education Today, 99*, 104815. https://doi.org/10.1016/j.nedt.2021.104815

Rudolph, J. W., Simon, R., Rivard, P., Dufresne, R. L., & Raemer, D. B. (2007). Debriefing with good judgment: combining rigorous feedback with genuine inquiry. *Anesthesiol Clinics, 25*(2), 361-376. https://doi.org/10.1016/j.anclin.2007.03.007

Schlairet, M. C., Schlairet, T. J., Sauls, D. H., & Bellflowers, L. (2015). Cognitive load, emotion, and performance in high-fidelity simulation among beginning nursing students: a pilot study. *Journal of Nursing Education, 54*(3 Suppl), S5-11. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med12&AN=25692940

Sewell, J. L., Maggio, L. A., Ten Cate, O., van Gog, T., Young, J. Q., & O’Sullivan, P. S. (2019). Cognitive load theory for training health professionals in the workplace: A BEME review of
studies among diverse professions: BEME Guide No. 53. *Medical Teacher, 41*(3), 256-270. https://doi.org/10.1080/0142159x.2018.1505034

Stefanidis, D., Scebo, M. W., Sechrist, C., Mostafavi, A., & Heniford, B. T. (2008). Do novices display automaticity during simulator training? *The American Journal of Surgery, 195*(2), 210-213. https://doi.org/10.1016/j.amjsurg.2007.08.055

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*(1), 1-16. https://doi.org/10.1007/s11423-019-09701-3

Szulewski, A., Howes, D., van Merriënboer, J. J. G., & Sweller, J. (2021). From Theory to Practice: The Application of Cognitive Load Theory to the Practice of Medicine. *Academic Medicine, 96*(1), 24-30. https://doi.org/10.1097/acm.0000000000003524

Tofil, N. M., Cheng, A., Lin, Y. Q., Davidson, J., Hunt, E. A., Chatfield, J., MacKinnon, L., & Kessler, D., &Int Network Simulation-Based, P. (2020). Effect of a Cardiopulmonary Resuscitation Coach on Workload During Pediatric Cardiopulmonary Arrest: A Multicenter, Simulation-Based Study. *Pediatric Critical Care Medicine, 21*(5), E274-E281. https://doi.org/10.1097/pcc.0000000000002275

Tremblay, M. L., Lafleur, A., Leppink, J., & Dolmans, D. H. (2017). The simulated clinical environment: Cognitive and emotional impact among undergraduates. *Medical Teacher, 39*(2), 181-187. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med14&AN=27832706

Tremblay, M. L., Leppink, J., Leclerc, G., Rethans, J. J., & Dolmans, D. H. J. M. (2019). Simulation-based education for novices: complex learning tasks promote reflective practice. *Medical Education, 53*(4), 380-389. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emexa&AN=626885850

Tufanaru, C., Munn, Z., Aromataris, E., Campbell, J., & Hopp, L. (2017). Chapter 3: Systematic reviews of effectiveness. In E. Aromataris & Z. Munn (Eds.), *Joanna Briggs Institute Reviewer’s Manual*. The Joanna Briggs Institute. https://reviewersmanual.joannabriggs.org/

Tyerman, J., Luctkar-Flude, M., Graham, L., Coffey, S., & Olsen-Lynch, E. (2019). A Systematic Review of Health Care Presimulation Preparation and Briefing Effectiveness. *Clinical Simulation in Nursing, 27*, 12-25. https://doi.org/10.1016/j.ecns.2018.11.002

Weigl, M., Stefan, P., Abhari, K., Wucherer, P., Fallavollita, P., Lazarovici, M., Weidert, S., Euler, E., & Catchpole, K. (2016). Intra-operative disruptions, surgeon’s mental workload, and technical performance in a full-scale simulated procedure. *Surgical Endoscopy, 30*(2), 559-566. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med13&AN=26091986

Wilson, M. R., Poolton, J. M., Malhotra, N., Ngo, K., Bright, E., & Masters, R. S. W. (2011). Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World journal of surgery, 35*(9), 1961-1969. https://doi.org/10.1007/s00268-011-1141-4

Wucherer, P., Stefan, P., Abhari, K., Fallavollita, P., Weigl, M., Lazarovici, M., Winkler, A., Weidert, S., Peters, T., de Ribaupierre, S., Eagleson, R., & Navab, N. (2015). Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis. *IEEE Transactions on Medical Imaging, 34*(8), 1730-1737. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med12&AN=25585414
Young, J. Q., Van Merrienboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Medical Teacher, 36*(5), 371-384. https://doi.org/10.3109/0142159x.2014.889290

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simulation in Healthcare, 5*(5), 267-271. https://doi.org/10.1097/SIH.0b013e3181e3f329

Zheng, R., & Greenberg, K. (2018). The Boundary Of Different Approaches In Cognitive Load Measurement : Strengths and Limitations. In R. Zheng (Ed.), *Cognitive Load Measurement and Application*. Routledge. https://doi.org/10.4324/9781315296258