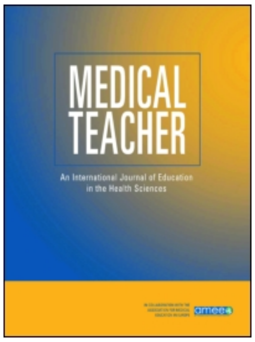

Medical Teacher

Taylor & Francis
Taylor & Francis Group

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: http://www.tandfonline.com/loi/imte20

Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions: BEME Guide No. 53

Justin L. Sewell, Lauren A. Maggio, Olle ten Cate, Tamara van Gog, John Q. Young & Patricia S. O'Sullivan

To cite this article: Justin L. Sewell, Lauren A. Maggio, Olle ten Cate, Tamara van Gog, John Q. Young & Patricia S. O'Sullivan (2018): Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions: BEME Guide No. 53 , Medical Teacher, DOI: 10.1080/0142159X.2018.1505034

To link to this article: https://doi.org/10.1080/0142159X.2018.1505034

Published online: 17 Oct 2018.

MEDICAL TEACHER
https://doi.org/10.1080/0142159X.2018.1505034

MEDICAL TEACHER

Taylor & Francis
Taylor & Francis Group

BEME GUIDE

Check for updates

Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions: BEME Guide No. 53

Justin L. Sewell[a], Lauren A. Maggio[b], Olle ten Cate[c,d], Tamara van Gog[e], John Q. Young[f] and Patricia S. O'Sullivan[d]

[a]Department of Medicine, Division of Gastroenterology, University of California San Francisco, San Francisco, CA, USA; [b]Department of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; [c]Center for Research and Development of Education, University Medical Center Utrecht, Utrecht, Netherlands; [d]Department of Medicine, Research and Development in Medical Education, University of California San Francisco, San Francisco, CA, USA; [e]Department of Education, Utrecht University, The Netherlands; [f]Department of Psychiatry, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA

ABSTRACT

Aim: Cognitive load theory (CLT) is of increasing interest to health professions education researchers. CLT has intuitive applicability to workplace settings, yet how CLT should inform teaching, learning, and research in health professions workplaces is unclear.

Method: To map the existing literature, we performed a scoping review of studies involving cognitive load, mental effort and/or mental workload in professional workplace settings within and outside of the health professions. We included actual and simulated workplaces and workplace tasks.

Result: Searching eight databases, we identified 4571 citations, of which 116 met inclusion criteria. Studies were most often quantitative. Methods to measure cognitive load included psychometric, physiologic, and secondary task approaches. Few covariates of cognitive load or performance were studied. Overall cognitive load and intrinsic load were consistently negatively associated with the level of experience and performance. Studies consistently found distractions and other aspects of workplace environments as contributing to extraneous load. Studies outside the health professions documented similar findings to those within the health professions, supporting relevance of CLT to workplace learning.

Conclusion: The authors discuss implications for workplace teaching, curricular design, learning environment, and metacognition. To advance workplace learning, the authors suggest future CLT research should address higher-level questions and integrate other learning frameworks.

Introduction

Researchers are increasingly interested in applying cognitive load theory (CLT) to health professions education (HPE) because of its potential to directly impact design of instruction (van Merrienboer and Sweller 2010). CLT focuses on the role of working memory in learning, developed by Sweller et al. in 1988 (Sweller 1988). As opposed to capacity for sensory input and long-term memory, both of which are theoretically unlimited, working memory is limited in both capacity and duration, allowing for maintaining and processing only a few pieces of information at any given time (Young et al. 2014). As a result, working memory functions as the primary bottleneck for learning. CLT posits that, during a learning task, the degree of working memory load (imposed by cognitive processes) should not exceed its limited capacity and that working memory should be used for processes conducive to learning rather than extraneous processes; otherwise, learning and performance will degrade (Sweller 1988; Young et al. 2014).

CLT envisions at least two, and possibly three, types of cognitive load (CL). *Intrinsic load* refers to cognitive demands imposed by accomplishing the essential components of a learning task. Intrinsic load owes largely to the complexity of the learning task and the learner's prior experience and knowledge. Ideally, intrinsic load should be

Practice points

- Studies of cognitive load in health and non-health professional workplace settings support relevance of cognitive load theory to workplace education and provide practical implications for ways to design workplace curricula, teach within the workplace, and organize workplace environments to optimize learners' cognitive load.
- Very high levels of cognitive load negatively impact health professions trainees' performance (and likely learning as well) in workplace settings. Novice learners and complex tasks and settings, particularly predispose to high levels of cognitive load.
- Aspects of workplace environments contribute to extraneous load, and can negatively impact capacity for engaging in activities that promote germane load and learning.
- Measuring cognitive load, particularly cognitive load subtypes, remains a primary challenge to fully understanding implications of cognitive load theory for health professions education workplaces.
- Utility of cognitive load theory for workplace learning may be enhanced by considering it within the context of other learning theories and frameworks.

CONTACT Justin L. Sewell justin.sewell@ucsf.edu Division of Gastroenterology, University of California San Francisco, Zuckerberg San Francisco General Hospital, 1001 Potrero Avenue, Unit 3D, San Francisco, CA 94110, USA

Supplemental data for this article can be accessed here.

matched to the learner's prior experience; that is, intrinsic load should be neither too high (which will overwhelm a learner's working memory) nor too low (which may induce boredom or apathy). *Extraneous load* occurs when learners use working memory resources to engage in cognitive processes that are not essential to completing the learning task. Common examples include external distractions (e.g. environmental) or suboptimal instructional design (e.g. unnecessarily having to search for information). Internal distractions (e.g. worries about external or personal issues, competing demands, self-induced time pressure) may also contribute to extraneous load (Feldon 2007; Young and Sewell 2015). Extraneous load should always be minimized. A third type of cognitive load, *germane load*, occurs when learners deliberately use cognitive processes to create or modify cognitive schemas (organized patterns of information held in long-term memory that are retrievable as a single unit). Examples of means to promote germane load include instructional design (e.g. interleaved practice compared to blocked practice) or prompting generative processes (e.g. self-explaining or elaborating) (Fiorella and Mayer 2016). Researchers currently debate whether to conceptualize germane load as a third type of cognitive load distinct from intrinsic and extraneous load or as a subset of intrinsic load (van Merrienboer and Sweller 2010; Kalyuga 2011; Leppink and van den Heuvel 2015; Young and Sewell 2015). For consistency, this paper will refer to germane load as a separate construct, but it can equally be conceptualized as "working memory resources used to deal with intrinsic load" (van Merrienboer and Sweller 2010; Kalyuga 2011; Leppink and van den Heuvel 2015). In either case, the ability to control or modulate germane load lies primarily with the learner, as opposed to intrinsic load, which is not typically under control of the learner (Young and Sewell 2015). Since elevated levels of intrinsic and extraneous load both reduce space for schema formation, elevated levels of either can negatively impact performance and learning.

An ideal learning task is one in which intrinsic is matched to the level of the learner, germane load is optimized, and extraneous load is minimized (Leppink et al. 2013; Young and Sewell 2015). Mental effort (ME) and mental workload (MWL) are concepts that predate CLT yet are related to CL, and are considered indicators of CL experienced by the learner (Paas et al. 2003). For consistency and ease of reading, we will use the term "CL" when referring generally to the related constructs of CL, ME, and MWL, but will use specific terms during discussion of studies that specifically refer to each construct.

CLT is particularly relevant to complex learning settings, such as found in HPE, where there are high levels of element interactivity (i.e. when elements of a task cannot be processed independent of one another but need to be processed in relation to each other for learning to occur) (van Merrienboer and Sweller 2010; Young et al. 2014; Fraser KL et al. 2015). Within HPE, workplace settings (both simulated and actual workplaces) are arguably among the most complex, often involving multiple tasks, numerous stakeholders, and prevalent distractions, in addition to urgent, emergent, and crisis situations in which high-stakes decisions require rapid, accurate responses. This workplace climate creates substantial potential for cognitive overload among learners.

Several CLT scholars have written excellent reviews of CLT in relation to HPE (van Merrienboer and Sweller 2010; Young et al. 2014; Fraser et al. 2015; Naismith and Cavalcanti 2015). However, substantial coverage of CLT's role in workplace learning is limited. Naismith and Cavalcanti's systematic review of validity evidence for CL measures was limited to simulation-based learning and did not include other workplace settings (Naismith and Cavalcanti 2015). Consequently, how and when CLT can be most effectively used to understand and guide instruction and research in HPE workplaces is unclear.

CLT has most often been discussed in relation to design of instruction within classroom learning settings, yet we appreciate strong potential for applications to workplace learning. We, therefore, designed a scoping review to map the existing literature related to CLT and its related constructs within workplace learning settings (both simulated and actual workplaces). Since study of non-healthcare settings has informed best practices in healthcare (e.g. insights from aviation have shaped the practice of anesthesia (Toff 2010)), we searched for studies in professional workplace settings both within and outside of the health professions. We were interested in both theoretical and practical implications of published studies. We designed the review to address three *a priori* research questions:

1. How do studies of CLT, CL, ME, and MWL in workplace settings inform, contribute to, or conflict with, theoretical tenets of CLT?
2. What practical implications for workplace teaching, curricular design, and educational research in the health professions can be drawn from included studies?
3. How has the study of CLT differed in health professions versus non-health professions settings, and what lessons can be learned from these differences?

Methods

We designed a scoping review following the six-step process described by Arksey and O'Malley (Arksey and O'Malley 2005). The Best Evidence in Medical Education (BEME) Collaboration (Thistlethwaite and Hammick 2010) approved the protocol that guided our methods. The first step – *Identifying the Research Question* – was covered in the Introduction. Steps 2–6 are described below.

Identifying relevant studies

We designed our search strategy to reflect two primary constructs: CL and workplace learning (including both simulated and actual workplaces and tasks). We searched eight databases, several of which include gray literature, to maximize representation of both HPE and non-HPE settings (see Supplemental Table A). Search strings were developed amongst authors, including those with experience in CLT (JLS, TvG, JQY), workplace learning (OtC, PSO'S), and library and information science (LAM). We slightly modified preliminary search terms after checking results of two pilot searches for 11 papers related to CLT that we expected our search terms should have produced (Khalil et al. 2008; van Merrienboer and Sweller 2010; Qiao et al. 2014; Young

Table 1. Study inclusion criteria.

Inclusion criterion	Definition/explanation
Study uses CLT or measures cognitive load	Study design explicitly uses CLT, *OR* measures cognitive load, mental effort, or mental workload.
Educational research study	Empirical research study that includes practitioners, learners, trainees, and/or student participants with specific study objective to assess process or outcomes of learning.
Workplace setting	Study occurs in an applied setting in which trainees or learners are involved in activities that are partial or complete representations of tasks and/or environments in which their profession is/will be applied.

et al. 2014; Andersen et al. 2016b; Chen et al. 2015a; Fraser et al. 2015; Haji, Rojas, et al. 2015; Naismith et al. 2015; Szulewski et al. 2015; Rasmussen et al. 2016). This resulted in adding terms related to learners, trainees, and students, after which all 11 studies were identified. Supplemental Table A depicts databases and search terms. De-duplication was performed initially with EndNote and later by hand.

We hand searched bibliographies of included studies and several review articles (van Merrienboer and Sweller 2010; Young et al. 2014; Fraser KL et al. 2015; Naismith and Cavalcanti 2015), and continually monitored the literature by regularly reviewing newly published tables of contents for several HPE journals (including *Medical Education, Academic Medicine, Medical Teacher, Advances in Health Sciences Education,* and *Teaching and Learning in Medicine*) for relevant articles published after the literature search was performed. Because of the very large number of journals that could have published relevant papers over a large timeframe, we did not feel that hand searching tables of contents of previously published journal issues was feasible. No restrictions were imposed on publication date, language, study design, or publication type.

The literature search was initially performed on 14 March 2016 and rerun 21 July 2017.

Selecting studies to be included in the review

Table 1 describes inclusion criteria; all three criteria were required for inclusion. Based on the second criterion, we included only empirical research and excluded reviews, commentaries, and editorials.

Review for inclusion or exclusion was accomplished using the Covidence platform (Veritas Health Innovation Ltd, Melbourne, Australia), a web-based systematic review manager. Two authors (JLS and PSO'S) independently reviewed titles and abstracts. They met after the first 25 titles to ensure that they consistently interpreted the inclusion criteria. They met periodically to work out disagreements through discussion. Full text screening was then conducted simultaneously, with JLS and PSO'S meeting multiple times to discuss studies marked for full-text review, coming to consensus on each for inclusion or exclusion. LAM was available at all stages to assist in decision-making if JLS and PSO'S could not agree on inclusion/exclusion.

Charting the data

We extracted pre-determined data that would inform our research questions: publication characteristics, profession, and training level of subjects, sample size, methodology, method of CL measurement, study outcomes, degree of CLT integration based on Kumasi's method (Kumasi et al. 2013), theoretical implications for CLT, and practical implications for workplace teaching, curriculum development, and research.

As the primary author, JLS extracted data for all studies, and second extractor duties were divided among other authors. Differences in data extraction were adjudicated through discussions between JLS and the other extractors. In the case that JLS and the other extractor could not agree on a particular piece of data, JLS discussed with PSO'S and/or LAM to resolve final coding decision. The authors extracted data using an online form created using the Qualtrics platform (Qualtrics, Provo, UT). The Supplemental Material depicts the data extraction form.

Collating, summarizing, and reporting results

JLS exported data from Qualtrics to an Excel spreadsheet. Characteristics of included studies were synthesized and reported in narrative and tabular format. Knowledge synthesis was performed as follows. Each data extractor coded "implications for workplace learning" and "implications for CLT" based on review of each study. JLS iteratively reviewed these comments within the Excel spreadsheet to develop an initial synthesized set of topics relevant to the research questions. These topics were refined and revised through discussion with all authors. Our pre-determined plan was to organize results, when possible, according to CLT's aims to optimize intrinsic load, minimize extraneous load, and optimize germane load (Leppink et al. 2013; Young and Sewell 2015), but we also planned to identify and consider themes that might emerge from the studies. We selected exemplar studies to highlight prominent findings. Selected data extracted from each study were also assembled into a table (Supplemental Table B) for further reference.

Since we sought to deliver a product that was not only theoretically relevant, but also practically useful, we set out to develop a set of "best practices" for applying CLT to workplace learning. To develop these best practices, we used the frameworks of CLT and workplace learning as lenses for synthesizing themes present in included studies. JLS thematically analyzed the "implications for workplace learning and CLT" that authors had coded through data extraction, as described above. JLS then synthesized these themes as they related to CLT and workplace learning, and assembled them into a set of "best practices," which was presented in tabular format. Based on our synthesis of included studies, all authors agreed upon the domains of curricular development, direct teaching, learning environment, and metacognition as a practical framework for organizing best practices. Best practices were edited iteratively until all authors agreed on their content and organization. Specific evidence supporting each best practice was cited.

JLS wrote the first draft of the manuscript following which all authors edited for key content.

Undertaking consultations with key stakeholders

On account of the diverse expertise among our authors, the qualifications of the BEME protocol reviewers, and the anticipated expertise of the editors and scholars, who would review our manuscript, we did not separately consult with additional stakeholders. The experiential knowledge of our team was primarily leveraged during design of the literature search (e.g. making sure that appropriate studies were identified) and through data analysis and knowledge synthesis (e.g. deciding how to organize and present studies and themes identified).

Results

Search results

Our searches retrieved 4571 citations, of which 352 full-texts were assessed for eligibility, with 116 studies included. Figure 1 shows the study selection flow. The Supplemental Material includes a Table B depicting all included studies.

Overview of included studies

Most studies included authors from North America ($N=82$) and/or Europe ($N=38$). The majority of studies ($N=103$) were published in the past decade, with an increasing publication trend over recent years. Nearly 30% of studies were published in 2016 or 2017.

Of the included studies, 97 were in HPE, mostly in medicine ($N=85$), but also nursing ($N=10$), surgical assisting ($N=3$), pharmacy ($N=2$), dentistry ($N=1$), and speech-language pathology ($N=1$). Aviation ($N=9$) was the most common non-HPE setting; other non-HPE professions included ship navigation, teachers, electricians, military, law enforcement, and nuclear power plant operators.

The majority of studies ($N=91$) were of simulation, whereas 24 occurred in non-simulated workplace settings. Most studies employed quantitative approaches that were experimental ($N=55$) or non-experimental ($N=49$). Few studies used qualitative ($N=5$) or mixed methods ($N=7$) designs. Median sample size was 27 subjects. See Table 2 for additional characteristics of included studies.

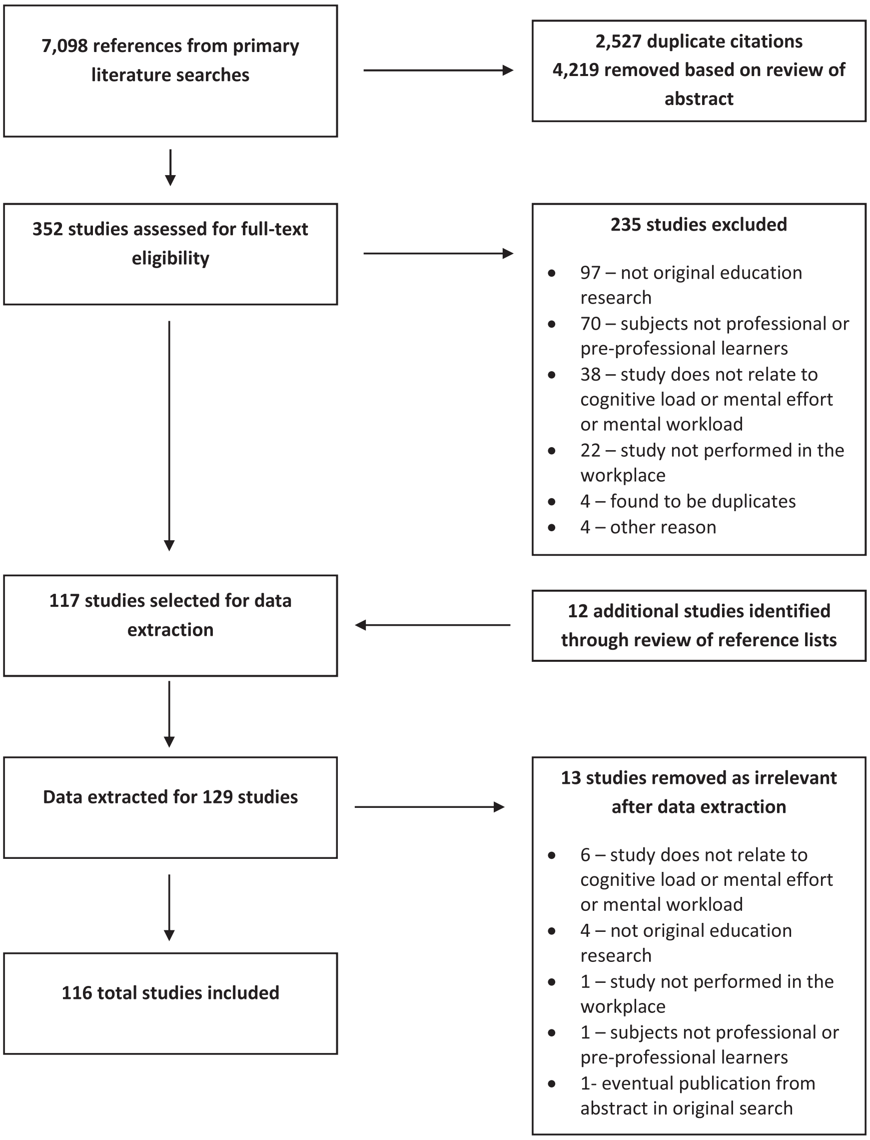

Figure 1. Study selection flow.

Table 2. Characteristics of included studies.

	Total $N = 116$
Continent of publication, No. (%)[a]	
North America	82 (70.7)
Europe	38 (32.8)
Asia	12 (10.3)
Australia	4 (3.4)
Africa	2 (1.7)
Study of health professions education, No. (%)	97 (83.6)
Study of non-health professions education, No. (%)	19 (16.4)
Health professions studied, No. (% in all studies; % in health professions studies only)[a]	
Medicine	85 (73.3, 87.6)
Nursing	10 (8.6, 10.3)
Surgical assistants	3 (2.6, 3.1)
Pharmacy	2 (1.7, 2.1)
Dentistry	1 (0.9, 1.0)
Speech-language pathology	1 (0.9, 1.0)
Non-health professions studied, No. (% in all studies; % in non-health professions studies only)	
Aviation	9 (7.8, 47.4)
Ship piloting	3 (2.6, 15.8)
Military	3 (2.6, 15.8)
Teachers	2 (1.7, 10.5)
Law enforcement	1 (0.9, 5.3)
Nuclear power plant operators	1 (0.9, 5.3)
Level of learners, No. (%)[a]	
Pre-professional students	6 (5.2)
Professional students	50 (43.1)
Post-graduate trainees	53 (45.7)
Practicing professionals	49 (42.2)
Sample size, median (range)	27 (4–573)
Study setting, No. (%)	
Simulation	91 (78.4)
Workplace	24 (20.7)
Combined simulation and workplace	1 (0.9)
Methodology	
Quantitative, non-experimental	49 (42.2)
Quantitative, experimental	55 (47.4)
Qualitative	5 (4.3)
Mixed methods, non-experimental	5 (4.3)
Mixed methods, experimental	2 (1.7)
CLT-related concepts mentioned, No. (%)[a]	
Cognitive load	61 (52.6)
Mental effort	35 (30.2)
Mental workload	70 (60.3)
Intrinsic load	26 (22.4)
Extraneous load	28 (24.1)
Germane load	22 (19.0)
Study measured/estimated cognitive load, mental effort, or mental workload, No. (%)	103 (88.8)
Type of cognitive load measured, No. (%)[a]	
Overall cognitive load, mental effort, mental workload	95 (81.9)
Subtype(s) of cognitive load	5 (4.3)
Overall and subtypes of cognitive load	3 (2.6)
Method for measuring CL, No. (%)[a]	
Psychometric single item	21 (18.1)
Psychometric multiple item	63 (54.3)
Secondary task	24 (20.7)
Physiologic measures	19 (16.4)
Qualitative methods	1 (0.9)

[a]Because more than one categorization could apply to an individual study, percentages may not add up to 100%.

Conceptualization of cognitive load

CL was most often conceptualized and studied as overall CL, rather than as CL subtypes (Table 2). Studies more often mentioned the construct of MWL ($N = 70$) than CL ($N = 61$) or ME ($N = 35$). The purpose of CL measurement was usually to compare CL between or amongst groups. Few studies mentioned intrinsic, extraneous, or germane load ($N = 26$, 28, and 22 respectively), and fewer ($N = 8$) attempted to measure one or more CL subtypes. Nuances of CLT such as expertise reversal effect (Kalyuga et al. 2003) and element interactivity (Hanham et al. 2017) were rarely mentioned.

Measurement of cognitive load

The large majority of studies ($N = 103$) attempted to measure or estimate CL, using diverse methods (Table 2). Psychometric approaches included single-item scales, usually Paas' scale (Paas et al. 2003), and multi-item scales. NASA-TLX was the most common multi-item scale, but several investigator-developed scales were utilized as well (see Supplemental Table B for details). Non-self-reported measures included responses to secondary task (both response time and accuracy), and a variety of physiologic measures (including heart rate and respiratory rate variability, electroencephalogram, electromyography, eye tracking, pupil diameter, blink frequency, serum adrenaline levels, brain imaging, and skin conductance). Several studies measured CL using more than one method: five studies used more than one psychometric approach, nine used both physiologic and psychometric approaches, and eight used both psychometric approaches and responses to secondary tasks. No studies used both physiologic measures and responses to secondary tasks.

Variables studied

The most commonly measured variables in quantitative studies were CL and performance. The latter was measured using many different techniques, usually applied immediately, rarely delayed. In qualitative and mixed-methods studies, CL was sometimes directly measured and was estimated qualitatively by subjects in others. Level of training and task complexity were the two most oft-studied independent variables in both quantitative and qualitative studies. One study explored the interaction between case complexity and learner knowledge during simulated handoffs and found that learner knowledge was a more important mediator of CL and performance (Young, Ten Cate, et al. 2016). A few studies assessed other covariates of CL and performance, including fatigue (Bertram et al. 1990; 1992; Tomasko et al. 2012), disruptions and distractions (Weigl et al. 2015; Gardner et al. 2016; Weigl et al. 2016; Thomas et al. 2017; Sexton et al. 2018), emotion (Fraser K et al. 2012, 2014; Naismith et al. 2015; Hautz et al. 2017), the learning environment (Sewell et al. 2017; Tremblay et al. 2017), and working with international students (Attrill et al. 2016).

Optimizing intrinsic load and overall cognitive load

When studies measured or discussed overall CL, it was nearly always operationalized as relating to intrinsic load. Numerous studies confirmed inverse relationships between task complexity (i.e. element interactivity, which was assessed in studies as either overall CL or intrinsic load) and performance, with performance being measured by how accurately tasks were performed, and/or how many errors were committed. A few examples (among many) included cadaveric versus simulated mastoidectomy (Andersen et al. 2016a), laparoscopic gynecological surgical simulations of varying complexity (Bharathan et al. 2013), simple versus complex simulated cardiac auscultation (Chen et al. 2015a), and simulated flight (Morris and Leung 2006). Links between CL induced by task complexity and performance were most pronounced among novice learners. Though most of these studies simply described links between level of training and intrinsic load, a few specifically aimed to reduce or optimize task complexity to improve learning. For example, in a study of engineering graduate students training to be nuclear power plant operators, a decision support system reducing complexity was found to reduce MWL and errors, and increase accuracy of decision-making, during abnormal operating procedures (Hsieh et al. 2012). In a study of psychiatry resident clinic panels, a workload-balancing method generated less variation in complexity of patients assigned to each resident, thus balancing anticipated CL and theoretically reducing risk of cognitive overload among these early trainees (Young et al. 2010). In a study involving pilots, a heads-up display designed to reduce pilots' need to derive system states reduced CL and improved performance (Cummings and Smith 2009). Notably, no studies addressed whether CL in workplaces could be "too low".

Simulation most commonly was used to study intrinsic or overall CL among early or novice learners (i.e. learners new to the field, or who had little experience performing the task in question). Beyond documenting benefits of simulation for learning, studies supported use of lower fidelity and lower complexity simulation for early learners. For example, Haji linked reduced element interactivity with higher performance and lower CL in a study of simple versus complex simulated suturing (Haji, Rojas, et al. 2015). Chen found similar findings related to performance in a study of simulated cardiac auscultation (Chen et al. 2015b), as did Tremblay in a study of simulated medication dispensing and patient counseling among pharmacy students (Tremblay et al. 2017). Among more advanced learners with adequate prior knowledge, however, Dankbaar's research suggested that a more complex simulation inducing greater intrinsic load could also promote germane load (Dankbaar et al. 2016).

Multiple studies assessed how simulated surgical approaches impacted overall CL or intrinsic load: including robotic versus laparoscopic (Stefanidis et al. 2010; Lee et al. 2014; Moore et al. 2015), simulation versus animal model (Mouraviev et al. 2016), endoscopic versus laparoscopic versus open (Zheng et al. 2014), and standard versus single incision laparoscopic (Montero et al. 2011; Di Stasi et al. 2017; Scerbo et al. 2017). These studies typically found the more complex approach to be associated with higher CL and lower performance. These differences were seen among both faculty and trainees alike, but were more marked among subjects with less prior experience (Stefanidis et al. 2010; Montero et al. 2011; Zheng et al. 2012, 2014; Britt et al. 2015; Moore et al. 2015; Andersen et al. 2016a; Mouraviev et al. 2016; Di Stasi et al. 2017; Scerbo et al. 2017).

While the above studies involved whole tasks, several studies, including three performed in non-HPE settings, aimed to determine which parts or aspects of workplace learning induced the greatest amount of CL. Moos found varying levels of overall CL among eight different teaching tasks performed by pre-service teachers (Moos and Pitton 2014). Dahlstrom identified that different flight segments caused varying levels of MWL among trainee and practicing pilots (Dahlstrom and Nahlinder 2009). In a study of ship piloting, Murai found that CL was highest when crucial decisions were required (Murai et al. 2010). Similar findings were seen in HPE settings including administration of anesthesia (Gaba and Lee 1990; Weinger et al. 2000), clinical reasoning in primary care (Shachak et al. 2009), and creating dental molds (Walker and von Bergmann 2015). In the latter study, Walker identified which task components were most complex using cognitive task analysis (Walker and von Bergmann 2015).

Although Moos' study suggested that learners may be able to recognize cognitive overload (Moos and Pitton 2014), no studies specifically taught learners or instructors to recognize or manage overall or intrinsic cognitive overload. Furthermore, all studies were focused on very specific settings, whereas none investigated CL at the large-scale curricular or overall workplace level.

Minimizing extraneous load

Studies with implications for extraneous load often focused on specific elements of workplace teaching environments that could impact CL. Notably, most studies measured overall CL rather than extraneous load. However, most manipulated variables were assumed or inferred to be inducers of extraneous load. Most common were studies of how information is displayed, interpreted, and manipulated within workplaces. Two HPE studies found lower CL during interpretation of clinical information when data were presented graphically rather than numerically, including arterial blood gas data (Doig et al. 2011) and critical care patient data (Workman et al. 2007). Dominessy documented similar findings among helicopter pilots; in that study, the method of presenting tactical information impacted CL; CL was highest when displayed as text, mid-level when displayed as numeric, and lowest when displayed in graphical format (Dominessy et al. 1991). Two studies found that more intuitive formatting of clinical reminders (Saleem et al. 2007) and order sets (Avansino and Leu 2012) in electronic health records could reduce CL (again extraneous load was implied); in the former study, learnability was improved, and in the latter, performance was improved. A qualitative study of an electronic health record also suggested that a well-designed electronic health record could reduce extraneous load, in particular, by making patient histories and test results more easily accessible, and by eliminating challenges associated with reading handwritten documentation (Shachak et al. 2009).

Researchers also studied reduction of extraneous load by standardizing tasks. This was primarily accomplished through the use of checklists, which were found to reduce CL (extraneous load implied) and improve performance. In aviation maintenance, Liang et al. found that instructions administered to aircraft mechanics using an online maintenance assistance platform was associated with lower CL and better performance than instructions written on a paper work card (Liang et al. 2010). In two different studies, Sibbald et al. found that use of checklists during simulated cardiac auscultation (Sibbald, de Bruin, Cavalcanti, et al. 2013) and simulated electrocardiogram interpretation (Sibbald et al. 2013) was associated with better performance, and equal or lesser CL, compared with a control condition.

In some studies, the physical environment was treated as an important variable, particularly as it impacted extraneous load. High fidelity environments were found to be associated with higher levels of extraneous load in simulated medication dispensing among pharmacy students, presumably due to increased prevalence of distractions within more authentic environments (Tremblay et al. 2017). In one study, the contextual factors of patient emotional volatility and non-English speaking increased CL and impaired performance (even among practicing physicians), presumably due to high levels of extraneous load (Durning SJ et al. 2012). Multi-tasking and time pressure increased CL and impaired nursing students' operation of infusion pumps (Kataoka et al. 2011) and surgical residents' laparoscopic suturing (Modi et al. 2016). Distractions and disruptions were found to be common sources of extraneous load during medication administration among nurses (Thomas et al. 2017) and among operating room staff (Weigl et al. 2015; 2016). Lee et al. completely redesigned the physical environment of a radiology group based on tenets of CLT, resulting in fewer disruptions, lower CL, and greater workplace satisfaction among fellows and faculty (Lee et al. 2017).

While the above focused on the physical environment, other studies focused on emotion as a potential internal contributor to extraneous load. Fraser studied emotions as a potential source of extraneous load in two studies. Among medical students participating in a simulation of aspirin overdose, half were randomized to a scenario in which the patient unexpectedly died. These subjects reported more negative emotions and higher CL (presumably extraneous load), and their performance three months later was statistically lower than the control group (Fraser et al. 2014). In the second study of medical students learning cardiac murmurs, invigoration was positively associated, and tranquility negatively associated with CL (again, presumably extraneous load), and CL was negatively associated with performance (Fraser et al. 2012).

Two studies documented teaching behaviors that could impact extraneous load. Tangential conversations were found to increase CL and impair performance during simulated laparoscopic suturing (Gardner et al. 2016). Greater teacher confidence and engagement were associated with lower levels of extraneous load among fellows learning to perform colonoscopy in actual workplace settings, though performance was not assessed (Sewell et al. 2017). In that same study, fatigue and number of prior procedures performed that day were positively associated with extraneous load.

Optimizing germane load

Optimizing intrinsic load and minimizing extraneous load creates space in working memory for activities contributing to germane load (Young et al. 2014). In addition, instructional and curricular approaches may promote germane load. There were relatively few studies of germane load, either explicitly stated or implied, despite optimal germane load being a primary goal of CLT. Since learning complex skills takes time (Zheng et al. 2010; Mohamed et al. 2014; Haji, Khan, et al. 2015; Haji, Rojas, et al. 2015; Haji et al. 2016), design of practice is important to promote germane load. Studies comparing distributed or spaced practice with massed practice (for simulated mastoidectomy and heart sound learning) produced contradictory results (Donato et al. 2014; Andersen et al. 2016b; Andersen, Konge, et al. 2016). Two studies of mixed and random practice (for simulated laparoscopic tasks and cardiorespiratory auscultation) found improved performance as compared with blocked practice (Chen et al. 2015a; Shewokis et al. 2017), which may imply greater learning and therefore germane load.

Studies suggested several specific teaching approaches that may promote germane load. Two studies found increased germane load or perception of learning when learners with adequate prior experience were taught using higher fidelity simulation: in simulated medication dispensing and patient counseling among pharmacy students (Tremblay et al. 2017), and in simulated emergency medical care among medical students (Dankbaar et al. 2016). Notably, in the former study, extraneous load was also higher in the high fidelity group. Other suggested methods to promote germane load included situational awareness training (Saus et al. 2006, 2010); self-explanation, asking clarifying questions, and/or confirming one's understanding (Young, Boscardin, et al. 2016); greater teacher engagement with learners (Sewell et al. 2017); and careful design of feedback practices (Lee and Lee 2018).

No studies specifically addressed how learners created or honed cognitive schemas. Only one study attempted to link germane load with future performance (Dankbaar et al. 2016). The authors did not find a link, possibly because the learning task (emergency medicine skills) was too complex for the learners' level of prior experience.

Crisis situations

Crisis situations emerged as a unique setting related to CL in workplace learning settings. Crisis situations may induce high levels of CL, not only because of the numerous informational elements and high element interactivity, but also because of time demands and stress-related emotions. Studies of crisis situations involved simulations of anesthesia (Davis et al. 2009), patient death (Fraser K et al. 2014), vertebroplasty (Wucherer et al. 2015), and postoperative care (Boet et al. 2017). Crisis situations tended to induce CL and reduce performance. In one study, debriefing after a crisis was associated with reduced CL during a subsequent simulated crisis as compared with no debriefing (Boet et al. 2017).

Comparing HPE and non-HPE studies

Non-HPE studies more often mentioned MWL (84.2% of non-HPE studies, 60.3% of HPE studies) and less often mentioned CL (21.1% of non-HPE studies, 52.6% of HPE studies) or a CL subtype (10.5% of non-HPE studies, 25.0% of HPE studies). Study designs, settings, and samples were not substantively different. Thematic analysis of non-HPE and HPE studies produced similar results with regard to theoretical considerations.

Following is a summary of major professions and topics present in non-HPE studies, noting that several were previously mentioned above. Many non-HPE studies evaluated operation of large vehicles (aircraft, helicopters, and ships). The aims of these studies were diverse and included: descriptively comparing CL and performance by level of training (Crosby and Parkinson 1979); testing whether cockpit innovations (such as method or formatting of information display) impacted CL or performance (Dominessy et al. 1991; Cummings and Smith 2009); comparing simulated and actual operation of the vehicles (Dahlstrom and Nahlinder 2009); determining the relative levels of CL during particular flight or navigation segments (Murai et al. 2010; Hsu et al. 2015); and determining whether increased CL was associated with increased errors (Morris and Leung 2006). Two studies of military and police personnel examined whether format of training (virtual, live, or mixed) (Maxwell and Zheng 2017), or additional training (situational awareness training) (Saus et al. 2006) impacted CL and performance. Two studies among pre-service teachers used mixed-methods approaches to assess the time course and impact of CL during teacher training (Broyles et al. 2011; Moos and Pitton 2014).

Discussion

In this scoping review, we identified 116 studies that investigated CLT, CL, MWL, and/or ME within simulated and actual workplaces among diverse professions. We performed our review in the context of other reviews of CLT and its relevance to HPE. In 2010, van Merrienboer and Sweller provided a focused review of CLT as a theory, and specific means to optimize CL in HPE (van Merrienboer and Sweller 2010). In 2014, Young et al. (an author on this study) summarized CLT tenets as applied broadly to medical education (Young et al. 2014). In 2015, Fraser et al. provided a review focused on designing simulation-based learning (Fraser KL et al. 2015). In 2015, Naismith and Cavalcanti performed a systematic review assessing CL measurement methods in simulation-based medical education (Naismith and Cavalcanti 2015). Each of these reviews has provided invaluable insights, on which our review builds and expands, in three important ways. First, we employed a rigorous systematic literature search strategy, which was not reported in the above reviews other than Naismith and Cavalcanti. This improves likelihood of maximally identifying relevant studies. Second, we focused exclusively on workplace settings, both simulated and actual, to develop practical recommendations for using CLT to inform workplace-based learning. While the Fraser and Naismith reviews both addressed simulation – which our review included – they were specifically focused on design and CL measurement during simulations, whereas our review considers workplace-based learning much more broadly. Third, our consideration of non-HPE workplace settings provided additional evidence for the relevance of CLT to workplace learning.

The 116 included studies provided information to address our three research questions. We found that the large majority of studies supported the primary tenets of CLT, and also provided new theoretical considerations for CLT. The studies provided numerous practical recommendations for workplace teaching, curricular design, and educational research, some of which have been previously suggested by others, and some of which are new. Studies outside of HPE provided additional evidence supporting relevance of CLT to workplace learning and provided suggestions for future research. In this Discussion section, we first will focus on two major issues in CLT: CL subtypes and measurement of CL. We will then discuss theoretical considerations and implications for research more broadly. We will conclude with practical implications for curricular design and direct teaching in HPE workplaces.

Theoretical considerations and implications for research

Cognitive load subtypes

Theoretical discussion of CLT is enriched and made more practical when CL subtypes are considered, yet studies infrequently measured ($N = 8$ studies, 6.9%) or discussed ($N = 29$ studies, 25.0%) intrinsic, extraneous, or germane load. Study designs frequently inferred overall CL or MWL as primarily relating to intrinsic load, for which a traditional definition (inherent difficulty of the required components of a learning task) was universally supported. Accordingly, intrinsic load was typically positively associated with task complexity and negatively associated with prior experience and performance. Of particular relevance to HPE workplace learning was fatigue, which was associated with intrinsic load in a few studies (Bertram et al. 1990; 1992; Tomasko et al. 2012; Sewell et al. 2017). Fatigue was also associated with extraneous load in one study (Sewell et al. 2017). Regardless of which type of CL fatigue primarily affects, these findings support efforts to monitor and mitigate fatigue in HPE workplaces.

Overall, intrinsic load was largely studied in a descriptive manner, with few studies attempting to modulate intrinsic load in workplace settings. This approach, coupled with clear inverse links between CL and performance, and lack of measures of actual learning, could tacitly promote an overly simplistic notion that lower intrinsic load is better for learning. However, as the goal of CLT is to optimize, not minimize, intrinsic load, this assertion lacks support from CLT, except perhaps among very novice learners (i.e. when the whole task is too complex and overwhelming for learning to occur). If CLT is to better inform workplace learning in HPE, it is imperative that future studies assess how to best optimize intrinsic load in HPE workplaces, in particular, how to match intrinsic load to learners' prior experience and competence, as has been discussed in classroom settings (Bannert 2002), so that levels of intrinsic load are neither too high (which may cause cognitive overload) nor too low (in which case there may be "nothing to learn" which might induce boredom and apathy toward learning). Such adaptive instruction has shown benefit for training effectiveness in experimental non-workplace settings (Camp et al. 2001; Corbalan et al. 2008). Studies have further suggested that, with training, students may themselves be able to select learning tasks appropriate for their zone of proximal development (Kostons et al. 2012). Although potentially more challenging in workplace settings, these two concepts may be adaptable for some HPE workplaces and tested as means to optimize intrinsic load.

Studies also supported a traditional view of extraneous load (i.e. non-essential elements of a learning setting requiring attention and mental effort). As shown in classroom-based research, the design of tools and technology that learners interact within workplaces were found to contribute to extraneous load (Dominessy et al. 1991; Saleem et al. 2007; Workman et al. 2007; Shachak et al. 2009; Doig et al. 2011; Avansino and Leu 2012). One study also discussed the physical environment as a source of extraneous load (Tremblay et al. 2017). Distractions, disruptions, and interruptions were another common source of extraneous load studied (Moos and Pitton 2014; Weigl et al. 2015; Gardner et al. 2016; Weigl et al. 2016; Thomas et al. 2017). In addition, contextual factors, time pressure, and multitasking were found as contributing to extraneous load in workplaces (Gaba and Lee 1990; Durning S et al. 2011; Kataoka et al. 2011; Modi et al. 2016). One aspect of the physical environment not studied was music. Music is commonly played in operating theaters and other HPE workplaces, and could distract learners, or could block out other internal or external distractions, thereby augmenting attention and performance (Moris and Linos 2013). Indeed, some studies have found that music promotes performance in surgery among novices (Siu et al. 2010), while others identified negative impact on learners' performance (Miskovic et al. 2008). Notably, these studies did not address CL. This could be an interesting area for future CLT research. Several studies sought to modulate extraneous load in workplaces, and these were all met with success (Saleem et al. 2007; Liang et al. 2010; Avansino and Leu 2012; Sibbald, de Bruin, Cavalcanti, et al. 2013; Sibbald et al. 2013; Lee et al. 2017).

Since workplace environments are typically more difficult to control than classroom settings, changes to minimize sources of extraneous load on learners may not always be feasible. Furthermore, influences contributing to extraneous load are endemic in workplaces in which learners will eventually work. This contributes to a contemporary question related to CLT: can we, and should we, train learners to disregard, deprioritize, and/or manage extraneous load during workplace training (Young, Wachter, et al. 2016)? Studies in university students suggest that, with experience, students develop an ability to disregard irrelevant information (Rop et al. 2018). Considering the above-noted prevalence of extraneous load in HPE workplaces, it is possible that providing HPE learners with strategies to manage sources of extraneous load early in training could reduce frequency of cognitive overload and provide ongoing benefits by increasing resilience and potentially reducing risk of burnout (Rotenstein et al. 2016), in addition to freeing working memory space for activities promoting germane load.

Germane load remains the subject of some debate, specifically whether it is a unique construct or best conceptualized as a part of intrinsic load. Few studies specifically measured germane load; those that did more often supported germane load as a distinct construct (Dankbaar et al. 2016; Sewell et al. 2016; Young, Irby, et al. 2016; Sewell et al. 2017) than not (Young, Boscardin, et al. 2016). It is intuitive that learner effort and metacognitive skills will promote learning and that these factors are distinct from the actual intrinsic learning task. Importantly, activities contributing directly to germane load are under the control of the learner, whereas factors contributing to intrinsic load are not (Young and Sewell 2015). Additional emerging evidence supports separate identities for intrinsic and germane load (Cook et al. 2017). The studies included in this review do not answer the question of germane load's "true identity," which remains a fruitful area for research and scholarly dialog, and would be informed by more sophisticated methods for measuring CL subtypes. However, we assert that discussion related to germane load should augment, and not interfere with, efforts to study best practices for teaching and learning from the perspective of CLT.

Like all learning theories, CLT views learning through a specific lens. It focuses strongly on consequences of the limitations of the cognitive architecture of the individual, in relation to the learning task, for learning and performance. However, in complex HPE workplace learning settings, the influence of the task may be difficult to disentangle from sociocultural aspects of workplace learning environments, as compared with less complex laboratory or classroom settings. The difficulty inherent in designing studies within complex workplaces may explain, at least in part, why many studies we identified were experimental and reductionist, and rarely examined an overall learning environment, clinical rotation, or workplace curriculum. As such, studies did not clearly indicate whether focused attempts to lower extraneous load or match intrinsic load to a learner's prior experience within a complex workplace setting would likely lead to linear increases in capacity for germane load and learning. Simply reducing the complexity of a particular workplace task might expose learners to information or tasks for which their learning schemas are not yet prepared. For this reason, understanding how learners form schemas, and helping them to articulate their schemas may be critical to promote accommodative learning (Illeris 2009), in which learners must break down part of a learning schema to integrate new information. Since the workplace is more fluid and is less controlled than the classroom, studying these processes is complex but could be of significant benefit to inform CLT and workplace teaching. Such efforts might benefit from integrating multiple theoretical frameworks.

Measuring cognitive load

Measurement or estimation of CL is a major challenge that may limit broader study of CLT in education settings (Moreno 2010), particularly HPE. Multiple studies employed more than one method to measure CL or MWL, and often demonstrated at least some agreement among measurement methods. However, the broad heterogeneity of included studies prevents drawing conclusions about any particular CL measurement technique as superior.

Measuring CL in (non-simulated) workplaces is particularly challenging, because measurement methods must avoid significantly interfering with task performance. For this reason, psychometric approaches are appealing, yet these tend to be performed as a summative end-of-task measure and do not capture variation in CL throughout a task, which was shown to vary across components of workplace tasks in several studies. Furthermore, CL measures are most informative when they separately measure CL subtypes, which can currently only be assessed using multi-item psychometric instruments that invariably disrupt real-time workplace tasks. This suggests a need for unobtrusive methods, possibly physiological, to continuously monitor CL and its subtypes. Some physiologic measures (such as heart rate) can be measured unobtrusively, yet cannot differentiate between CL subtypes. Physiologic measurements as we currently understand them do not afford this possibility, but as technology continues to advance, this should be an area of priority for innovation. It is plausible that capabilities of wearable technology such as smart watches or electroencephalography headsets (Guru et al. 2015; Arico et al. 2016; Hussein et al. 2016) could one day be harnessed for this purpose. Until such measures are available, one option is carefully crafted studies in which CL subtypes are systematically manipulated and assessed in high-fidelity workplace simulations, followed by assessment of performance in actual workplaces. Another potential approach would be more nuanced studies (possibly with mixed methods) that systematically examine processes and outcomes of learning within actual workplaces, and link when possible to CL. Such studies might allow more direct conclusions to be drawn regarding CL and its subtypes in HPE workplaces.

Related to measurement of CL is the need for recognizing cognitive overload among learners. Numerous studies suggested that secondary task techniques may identify overloaded learners. The need for hardware and constant monitoring may limit their utility in some workplaces (as opposed to simulation), yet there may be some workplace tasks that would lend themselves to periodic secondary task assessment. It is plausible that experienced teachers could recognize when learners are overwhelmed, based on body language and utterances; this has not been empirically studied, and would be an interesting area for investigation. Likewise, training learners to recognize and act upon their own cognitive overload could be empowering.

Narrowing and broadening the scope of CLT research in HPE workplaces

Despite the broad variety of studies identified through our search, the vast majority were very narrowly focused on specific learning settings, most often simulated and experimental settings. Many of these studies used CLT or the concept of MWL to address a specific teaching question, whereas very few were designed to query CLT and/or advance the theory.

As the tenets of CLT are generally agreed upon and broadly supported by research – questions about germane load notwithstanding (Leppink and van den Heuvel 2015; Young and Sewell 2015) – we propose it is time to broaden the scope of CLT research in the workplace and to address more sophisticated and actionable gaps in the literature such as those highlighted in Box 1. We note that the ability to address these topics may be enhanced by considering impact of other learning frameworks and the complexity of HPE workplace settings.

Box 1. Gaps in the existing literature identified by our review.

- What workplace factors contribute to the different types of CL?
- How might very high or very low levels of CL impact workplace learning?
- How do emotion and mindset impact CL in HPE workplaces?
- How can workplace curricula optimize intrinsic load, minimize extraneous load, and optimize germane load among diverse groups of learners with differing levels of prior knowledge and experience?
- How might Inter-professional work affect CL in HPE workplaces?
- How can teachers identify and act upon cognitive overload in their workplace learners?
- How can learners identify and act upon their own cognitive overload?
- How does cognitive overload impact learner stress and burnout?

With increasing emphasis on competency-based medical education (Gruppen et al. 2017) and burgeoning interest in variable duration training (Lucey et al. 2018), it is increasingly imperative that learning among HPE trainees is optimized, particularly in settings simulating or authentically representing the workplaces in which they will eventually work (once adequate expertise has been attained). We chose CLT as the theoretical framework through which to perform our scoping review, because we found CLT to be highly relevant to such workplace settings, as it provides a practical and pragmatic approach through which to consider design of individual learning sessions and to some extent overall curricula. However, CLT remains a theory specifically focused on certain (not all) cognitive aspects of learning, and study designs were largely limited to narrowly defined experimental settings, short-term outcomes, and lacking clear evidence for transfer to more authentic workplace settings. In addition, few studies mentioned learning theories or frameworks apart from CLT. Our review and synthesis of included studies suggests that HPE workplace research might benefit from studies integrating CLT with other cognitive (e.g. encapsulation theory) and "non-cognitive" (e.g. sociocultural learning, motivation) theories of workplace learning (Lane 2010). We suspect that perspectives offered by different theories could more fully capture the complexity of HPE workplace learning.

The studies in our review also provide guidance as to several questions that do *not* require further study. Links between intrinsic load (or overall CL or MWL) and prior experience, task complexity, and immediate performance are incontrovertible. It is evident that CL is lower, and performance higher, among more versus less knowledgeable learners, in lower versus higher complexity simulations, and in simulated tasks compared with actual workplace tasks. It is also clear that novice or early learners should start with tasks of lower complexity, lower fidelity, and less authenticity (to reduce risk of cognitive overload from very high intrinsic load), and that workplace learning curricular should take into account the individual learner's prior experience and current competence (and ideally match intrinsic load to the level of the learner).

Further descriptive study of the forgoing relationships is unlikely to move the field forward. We propose that these be considered as established truth without need for ongoing investigation. However, though these relationships need not be studied descriptively and retrospectively, they should certainly be used to inform innovations that could be developed and tested prospectively. In other words, these findings should be translated into actual practice. For example, using the established relationship between these to create an adaptive sequence of training tasks, as has been attempted in classroom settings (Camp et al. 2001; Mihalca et al. 2011), could be of immense benefit to workplace training in HPE settings.

Non-HPE studies

We included studies from professions outside of healthcare to inform and enrich our study findings and implications. A primary finding was that, despite teaching cultures and practice settings that may differ markedly from the health professions, theoretical findings from HPE and non-HPE studies were well aligned. This provides additional evidence supporting the relevance of CLT to professional workplace education. We also found intuitive connections between certain professions within and outside the health professions. For example, the non-HPE profession of pilot appeared analogous to procedural fields within the health professions, such as surgery, whereas more cognitive fields outside the health professions, such as air traffic control (Arico et al. 2016) and nuclear power point operation (Hsieh et al. 2012), seemed relevant to cognitive medical fields in which large volumes of information are exchanged at high rates (e.g. pharmacy or critical care nursing). Finally, the two studies among pre-service teachers (Broyles et al. 2011; Moos and Pitton 2014) provided insights into processes likely experienced among health professionals as they learn to become health professions educators.

Practical recommendations for workplace teaching and curricular design

CLT provides guidance for how cognitive activities of the working memory should be allocated. For example, to optimize germane load (or manage intrinsic load), learners should be given learning tasks of appropriate complexity (i.e. match intrinsic load to the level of the learner) with instructions, teaching, and a learning environment conducive to learning with minimal distractions (i.e. minimize extraneous load). Since HPE workplaces involve care for real patients, these goals may be more difficult to accomplish in workplaces as compared with classroom settings. These challenges inspired this review, and the studies we identified provide insight into pedagogical approaches to managing CL in HPE workplaces. Notably, though most studies in our review did not specifically measure or in many cases mention subtypes of CL, the somewhat reductionist designs of many studies permitted extrapolation as to the type of CL being modified. Based on our synthesis of the literature, and our experiences with CLT and workplace learning, we developed several "best practices" (development process is detailed in Methods section). Through our synthesis, we decided to categorize these best practices and practical recommendations into those related to curricular design, direct teaching, learning environment, and metacognition in Supplemental Table 3. Evidence from included studies is provided when relevant. While some of these recommendations were supported in prior reviews of CLT (van Merrienboer and Sweller 2010; Young et al. 2014), we discuss them here as they impact workplace learning in HPE.

Limitations

We used a scoping methodology (Arksey and O'Malley 2005) for this review to address our research questions and accomplish our overall goal of mapping the literature describing study of CLT in professional workplace learning. Although a scoping review was the most appropriate methodology to accomplish our aims, the scoping approach has limitations, particularly lack of study quality assessment. Given the large number of fairly heterogeneous and largely descriptive studies, as well as overall consistency of findings related to CLT, we think it unlikely that formal study quality assessment would have substantively impacted our synthesis. However, we remind readers to consider the best practices we describe as well as other claims in this review in the context the scoping review methodology and its lack of study quality assessment.

Characteristics of our literature search could have reduced the likelihood of identifying particular studies, and it is possible that our search failed to identify some studies that could have been relevant. Such limitations are inherent in any review, and we believe that inclusion of numerous databases, use of broad search terms, and inclusion of all search dates and languages should have minimized missed studies. Publication bias (i.e. studies with positive results might be more likely published than 'negative' studies) could have impacted studies we identified and our resultant synthesis. Such studies might be more likely published as abstracts only, and some databases we searched do include conference abstracts.

We attempted to quantify degree of theory integration using an adapted version of Kumasi's theory integration scale (Kumasi et al. 2013), but this proved challenging for the large number of studies that included a measure of MWL but did not mention or cite CLT. Many of these studies included only a single reference supporting the method for MWL measurement, or no relevant citations at all. This might suggest a low level of theory integration, and yet, the integration of the construct of MWL throughout a study, including measurement, might imply a high level of theory integration. This caused numerous coding differences amongst coders that could not be systematically reconciled, despite multiple discussions. Despite the lack of quantitative data describing theory integration, we are comfortable making the overall statement that CLT and the constructs of CL, ME, and MWL were integrated at a high level in the majority of studies. Future researchers should consider how to objectively determine the degree of CLT integration in HPE research, and whether quantification provides practical benefits.

Summary

In this scoping review, we found strong support for established tenets of CLT, new information to be further studied, and guidance for applying CLT in HPE workplaces in the future. Issues related to measurement of CL remain and must be further investigated. We argue that CLT is highly applicable to HPE workplaces, yet it cannot fully explain the complexity of HPE workplaces. We propose that future research of CL in HPE workplaces would benefit from integration of additional learning theories and frameworks to further disentangle some of the sources of complexity inherent to authentic workplace settings.

Acknowledgments

We thank members of BEME and the BEME International Collaborating Centre for their expert input into protocol development and manuscript review.

We thank Holly Nishimura for her assistance in executing the literature searches and retrieving full-text articles.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Funding

The study was funded in part by an award from the Society of Directors of Research in Medical Education (SDRME). The funding source played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit for publication.

Glossary

1. **Cognitive load theory:** A theory of instructional design based on our knowledge of human cognitive architecture, in particular, the limitations of working memory (Sweller et al. 2011).

2. **Cognitive load:** Working memory demands imposed by performing a task (Young et al. 2014).

3. **Intrinsic load:** Cognitive load that occurs as learners accomplish the essential elements of a task (Young et al. 2014).

4. **Extraneous load:** Cognitive load that is not essential to completing the task (e.g. distractions) (Young et al. 2014).

5. **Germane load:** Cognitive load imposed by the learner's deliberate use of cognitive strategies to generate or refine learning schemas (Young et al. 2014).

6. **Mental effort:** The aspect of cognitive load related to cognitive capacity that is actually allocated to accommodate demands imposed by a task (Paas et al. 2003).

7. **Mental workload:** The aspect of cognitive load related to the interaction between characteristics of the learner and characteristics of the task (Paas et al. 2003).

Notes on contributors

Justin L. Sewell, MD, MPH, is an academic gastroenterologist and Associate Professor of Medicine at the University of California San Francisco (UCSF). Dr. Sewell is currently pursuing a PhD in Health Professions Education through a joint program of UCSF and Utrecht University in the Netherlands. His thesis work is focused on implications of cognitive load theory for workplace learning in the health professions.

Lauren A. Maggio, PhD, MS(LIS), is an Associate Professor of Medicine at the Uniformed Services University of the Health Sciences in Bethesda, MD. She has training in information science and holds a PhD in Health Professions Education from Utrecht University.

Olle ten Cate, PhD, is a Full Professor of Medical Education and Senior Scientist in the Center for Research and Development of Education at UMC Utrecht.

Tamara van Gog, PhD, is a Professor of Educational Sciences at Utrecht University with substantial expertise and experience in research related to cognitive load theory.

John Q. Young, MD, MPP, PhD, is an academic psychiatrist, Professor of Psychiatry, and the Vice Chair for Education, Zucker School of Medicine at Hofstra/Northwell in Hemstead, NY. He received a PhD in Health Professions Education from Utrecht University, and is an expert in cognitive load theory.

Patricia S. O'Sullivan, EdD, is a Professor of Medicine and the Director of Research and Development in Medical Education at UCSF.

ORCID

Justin L. Sewell http://orcid.org/0000-0003-4049-2874
Lauren A. Maggio http://orcid.org/0000-0002-2997-6133
Olle ten Cate http://orcid.org/0000-0002-6379-8780
John Q. Young http://orcid.org/0000-0003-2219-5657
Patricia S. O'Sullivan http://orcid.org/0000-0002-8706-4095

References

Andersen SA, Konge L, Caye-Thomasen P, Sorensen MS. 2016. Retention of mastoidectomy skills after virtual reality simulation training. JAMA Otolaryngol Head Neck Surg. 142:635–640.

Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. 2016a. Cognitive load in mastoidectomy skills training: virtual reality simulation and traditional dissection compared. J Surg Educ. 73:45–50.

Andersen SA, Mikkelsen PT, Konge L, Caye-Thomasen P, Sørensen MS. 2016b. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. Laryngoscope. 126:E74–E79.

Arico P, Borghini G, Di Flumeri G, Colosimo A, Bonelli S, Golfetti A, Pozzi S, Imbert JP, Granger G, Benhacene R. 2016. Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. Front Hum Neurosci. 10:539.

Arksey H, O'Malley L. 2005. Scoping studies: towards a methodological framework. Int J Soc Res Method. 8:19–32.

Attrill S, Lincoln M, McAllister S. 2016. Supervising international students in clinical placements: perceptions of experiences and factors influencing competency development. BMC Med Educ. 16:180

Avansino J, Leu MG. 2012. Effects of CPOE on provider cognitive workload: a randomized crossover trial. Pediatrics. 130:e547–e552.

Bannert M. 2002. Managing cognitive load–recent trends in cognitive load theory. Commentary. Learn Instr. 12:139–146.

Bertram DA, Hershey CO, Opila DA, Quirin O. 1990. A measure of physician mental work load in internal medicine ambulatory care clinics. Med Care. 28:458–467.

Bertram DA, Opila DA, Brown JL, Gallagher SJ, Schifeling RW, Snow IS, Hershey CO. 1992. Measuring physician mental workload: reliability and validity assessment of a brief instrument. Med Care. 30:95–104.

Bharathan R, Vali S, Setchell T, Miskry T, Darzi A, Aggarwal R. 2013. Psychomotor skills and cognitive load training on a virtual reality laparoscopic simulator for tubal surgery is effective. Eur J Obstet Gynecol Reprod Biol. 169:347–352.

Boet S, Sharma B, Pigford AA, Hladkowicz E, Rittenhouse N, Grantcharov T. 2017. Debriefing decreases mental workload in surgical crisis: a randomized controlled trial. Surgery. 161:1215–1220.

Britt RC, Scerbo MW, Montano M, Kennedy RA, Prytz E, Stefanidis D. 2015. Intracorporeal suturing: transfer from fundamentals of laparoscopic surgery to cadavers results in substantial increase in mental workload. Surgery. 158:1428–1433.

Broyles TW, Epler CM, Waknine JW. 2011. A case study examining the impact of cognitive load on reflection and pre-service teachers' transfer of specific teaching behaviors. Career and Technical Education Research. 36:49–66.

Camp G, Paas F, Rikers R, van Merrienboer J. 2001. Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. Comp Hum Behav. 17:575–595.

Chen HC, O'Sullivan P, Teherani A, Fogh S, Kobashi B, ten Cate O. 2015. Sequencing learning experiences to engage different level learners in the workplace: An interview study with excellent clinical teachers. Med Teach. 37:1090–1097.

Chen R, Grierson L, Norman G. 2015a. Manipulation of cognitive load variables and impact on auscultation test performance. Adv Health Sci Educ Theory Pract. 20:935–952.

Chen R, Grierson LE, Norman GR. 2015b. Evaluating the impact of high- and low-fidelity instruction in the development of auscultation skills . Med Educ. 49:276–285.

Cook DA, Castillo RM, Gas B, Artino AR Jr. 2017. Measuring achievement goal motivation, mindsets and cognitive load: validation of three instruments' scores. Med Educ. 51:1061–1074.

Corbalan G, Kester L, van Merriënboer JJG. 2008. Selecting learning tasks: effects of adaptation and shared control on learning efficiency and task involvement. Contemp Educ Psychol. 33:733–756.

Crosby JJ, Parkinson SR. 1979. A dual task investigation of pilots' skill level. Ergonomics. 22:1301–1313.

Cummings ML, Smith CA. 2009. Improving precision vertical flight operations through a direct-perception-action display. J Aircraft. 46:1270–1279.

Dahlstrom N, Nahlinder S. 2009. Mental workload in aircraft and simulator during basic civil aviation training. Int J Aviat Psychol. 19:309–325.

Dankbaar ME, Alsma J, Jansen EE, van Merrienboer JJ, van Saase JL, Schuit SC. 2016. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. Adv Health Sci Educ Theory Pract. 21:505–521.

Davis DHJ, Oliver M, Byrne AJ. 2009. A novel method of measuring the mental workload of anaesthetists during simulated practice. Br J Anaesth. 103:665–669.

Di Stasi LL, Díaz-Piedra C, Ruiz-Rabelo JF, Rieiro H, Sanchez Carrion JM, Catena A. 2017. Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: Gaze-based indices. Appl Ergon. 65:168–174.

Doig AK, Albert RW, Syroid ND, Moon S, Agutter JA. 2011. Graphical arterial blood gas visualization tool supports rapid and accurate data interpretation. Comput Inform Nurs. 29:204–260.

Dominessy ME, Lukas JH, Malkin FJ, Monty RA, Oatman LC. 1991. The effect of information display formats on helicopter pilots' target acquisition and flying performance. Mil Psychol. 3:163–176.

Donato AA, Kaliyadan AG, Wasser T. 2014. Self-directed study using MP3 players to improve auscultation proficiency of physicians: a randomized, controlled trial. J Contin Educ Health Prof. 34:131–138.

Durning S, Artino AR, Jr., Pangaro L, van der Vleuten CP, Schuwirth L. 2011. Context and clinical reasoning: understanding the perspective of the expert's voice. Med Educ. 45:927–938.

Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. 2012. The impact of selected contextual factors on experts' clinical reasoning performance (does context impact clinical reasoning performance in experts?) Adv Health Sci Educ Theory Pract. 17:65–79.

Ericsson KA. 2004. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 79:S70–S81.

Feldon DF. 2007. Cognitive load and classroom teaching: the double-edged sword of automaticity. Educ Psych. 42:123–137.

Fiorella L, Mayer RE. 2016. Eight ways to promote generative learning. Educ Psychol Rev. 28:717–741.

Fraser K, Huffman J, Ma I, Sobczak M, McIlwrick J, Wright B, McLaughlin K. 2014. The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. Chest. 145:958–963.

Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. 2012. Emotion, cognitive load and learning outcomes during simulation training. Med Educ. 46:1055–1062.

Fraser KL, Ayres P, Sweller J. 2015. Cognitive load theory for the design of medical simulations. Simul Healthc. 10:295–307.

Gaba DM, Lee T. 1990. Measuring the workload of the anesthesiologist. Anesth Analg. 71:354–361.

Gardner AK, Clanton J, Jabbour II, Scott L, Scott DJ, Russo MA. 2016. Impact of seductive details on the acquisition and transfer of laparoscopic suturing skills: emotionally interesting or cognitively taxing? Surgery. 160:580–585.

Gruppen L, Frank JR, Lockyer J, Ross S, Bould MD, Harris P, Bhanji F, Hodges BD, Snell L, Ten Cate O. 2017. Toward a research agenda for competency-based medical education. Med Teach. 39:623–630.

Guru KA, Esfahani ET, Raza SJ, Bhat R, Wang K, Hammond Y, Wilding G, Peabody JO, Chowriappa AJ. 2015. Cognitive skills assessment during robot-assisted surgery: separating the wheat from the chaff. BJU Int. 115:166–174.

Haji FA, Cheung JJ, Woods N, Regehr G, de Ribaupierre S, Dubrowski A. 2016. Thrive or overload? The effect of task complexity on novices' simulation-based learning. Med Educ. 50:955–968.

Haji FA, Khan R, Regehr G, Drake J, de Ribaupierre S, Dubrowski A. 2015. Measuring cognitive load during simulation-based psychomotor skills training: sensitivity of secondary-task performance and subjective ratings. Adv Health Sci Educ Theory Pract. 20:1237–1253.

Haji FA, Rojas D, Childs R, Ribaupierre S, Dubrowski A. 2015. Measuring cognitive load: performance, mental effort and simulation task complexity. Med Educ. 49:815–827.

Hanham J, Leahy W, Sweller J. 2017. Cognitive load theory, element interactivity, and the testing and reverse testing effects. Appl Cognit Psychol. 31:265–280.

Hautz WE, Schroder T, Dannenberg KA, Marz M, Holzer H, Ahlers O, Thomas A. 2017. Shame in medical education: a randomized study of the acquisition of intimate examination skills and its effect on subsequent performance. Teach Learn Med. 29:196–206.

Hsieh MH, Hwang SL, Liu KH, Liang SFM, Chuang CF. 2012. A decision support system for identifying abnormal operating procedures in a nuclear power plant. Nucl Eng Des. 249:413–418.

Hsu CK, Lin SC, Li WC. 2015. Visual movement and mental-workload for pilot performance assessment. Lect Notes Comput Sci. 9174:356–364.

Hussein AA, Shafiei SB, Sharif M, Esfahani E, Ahmad B, Kozlowski JD, Hashmi Z, Guru KA. 2016. Technical mentorship during robot-assisted surgery: a cognitive analysis? BJU Int. 118:429–436.

Illeris K. 2009. Contemporary theories of learning. New York, NY: Routledge.

Kalyuga S. 2011. Cognitive load theory: how many types of load does it really need? Educ Psychol Rev. 23:1–19.

Kalyuga S, Ayres P, Chandler P, Sweller J. 2003. The expertise reversal effect. Educational Psychologist. 38:23–31.

Kataoka J, Sasaki M, Kanda K. 2011. Effects of mental workload on nurses' visual behaviors during infusion pump operation. Jpn J Nurs Sci. 8:47–56.

Khalil MK, Paas F, Johnson TE, Su YK, Payer AF. 2008. Effects of instructional strategies using cross sections on the recognition of anatomical structures in correlated CT and MR images. Anat Sci Ed. 1:75–83.

Kostons D, van Gog T, Paas F. 2012. Training self-assessment and task-selection skills: a cognitive approach to improving self-regulated learning. Learn Instr. 22:121–132.

Kumasi KD, Charbonneau DH, Walster D. 2013. Theory talk in the library science scholarly literature: an exploratory analysis. Lib Inf Sci Res. 35:175–180.

Lane IF. 2010. Professional competencies in health sciences education: from multiple intelligences to the clinic floor. Adv Health Sci Educ Theory Pract. 15:129–146.

Lee GI, Lee MR. 2018. Can a virtual reality surgical simulation training provide a self-driven and mentor-free skills learning? Investigation of the practical influence of the performance metrics from the virtual reality robotic surgery simulator on the skill learning and associated cognitive workloads. Surg Endosc. 32:62–72.

Lee GI, Lee MR, Clanton T, Clanton T, Sutton E, Park AE, Marohn MR. 2014. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. Surg Endosc. 28:456–465.

Lee MH, Schemmel AJ, Pooler BD, Hanley T, Kennedy T, Field A, Wiegmann D, Yu JJ. 2017. Radiology workflow dynamics: how workflow patterns impact radiologist perceptions of workplace satisfaction. Acad Radiol. 24:483–487.

Leppink J, Duvivier R. 2016. Twelve tips for medical curriculum design from a cognitive load theory perspective. Med Teach. 38:669–674.

Leppink J, Paas F, Van der Vleuten CP, Van Gog T, Van MJJ. 2013. Development of an instrument for measuring different types of cognitive load. Behav Res Methods. 45:1058–1072.

Leppink J, van den Heuvel A. 2015. The evolution of cognitive load theory and its application to medical education. Perspect Med Educ. 4:119–127.

Liang GF, Lin JT, Hwang SL, Wang EM, Patterson P. 2010. Preventing human errors in aviation maintenance using an on-line maintenance assistance platform. Int J Ind Ergon. 40:356–367.

Lucey CR, Thibault GE, Ten Cate O. 2018. Competency-based, time-variable education in the health professions: crossroads. Acad Med. 93:S1–S5.

Maxwell D, Zheng J. 2017. Large scale testing and evaluation of virtual environments for infantry soldier tasks comparing mental effort for live versus virtual training performance assessments. Virginia Beach, VA: MODSIM World 2017.

Mayer RE. 2009. Multimedia learning. 2nd ed. New York, NY: Cambridge University Press.

Mihalca L, Salden RJCM, Corbalan G, Paas F, Miclea M. 2011. Effectiveness of cognitive-load based adaptive instruction in genetics education. Comp Hum Behav. 27:82–88.

Miskovic D, Rosenthal R, Zingg U, Oertli D, Metzger U, Jancke L. 2008. Randomized controlled trial investigating the effect of music on the virtual reality laparoscopic learning performance of novice surgeons. Surg Endosc. 22:2416–2420.

Modi H, Leff D, Singh H, Darzi A. 2016. The influence of cognitive load on technical ability among surgical trainees. Int J Surg. 36:S34–S35.

Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. 2014. Validation of the national aeronautics and space administration task load index as a tool to evaluate the learning curve for endoscopy training. Can J Gastroenterol Hepatol. 28:155–159.

Montero PN, Acker CE, Heniford BT, Stefanidis D. 2011. Single incision laparoscopic surgery (SILS) is associated with poorer performance and increased surgeon workload compared with standard laparoscopy. Am Surg. 77:73–77.

Moore LJ, Wilson MR, McGrath JS, Waine E, Masters RSW, Vine SJ. 2015. Surgeons' display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy. Surg Endosc. 29:2553–2560.

Moos DC, Pitton D. 2014. Student teacher challenges: using the cognitive load theory as an explanatory lens. Teaching Education. 25:127–141.

Moreno R. 2010. Cognitive load theory: more food for thought. Instr Sci. 38:135–141.

Moris DN, Linos D. 2013. Music meets surgery: two sides to the art of "healing". Surg Endosc. 27:719–723.

Morris CH, Leung YK. 2006. Pilot mental workload: how well do pilots really perform? Ergonomics. 49:1581–1596.

Mouraviev V, Klein M, Schommer E, Thiel DD, Samavedi S, Kumar A, Leveillee RJ, Thomas R, Pow-Sang JM, Su L-M, et al. 2016. Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules. J Robot Surg. 10:49–56.

Murai K, Ishikura A, Hayashi Y, Matsumoto T, Kuramoto K, Higuchi K, Saiki T, Fujita T, Maenaka K. 2010. Toward evaluation of ship bridge teamwork for marine simulator training based on heart rate variability. Kobe, Japan: World Automation Congress.

Naismith LM, Cavalcanti RB. 2015. Validity of cognitive load measures in simulation-based training: a systematic review. Acad Med. 90:S24–S35.

Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. 2015. Limitations of subjective cognitive load measures in simulation-based procedural training. Med Educ. 49:805–814.

Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. 2003. Cognitive load measurement as a means to advance cognitive load theory. Educ Psychol.38:63–71.

Qiao YQ, Shen J, Liang X, Ding S, Chen FY, Shao L, Zheng Q, Ran ZH. 2014. Using cognitive theory to facilitate medical education. BMC Med Educ. 14:79.

Rasmussen SR, Konge L, Mikkelsen PT, Sørensen MS, Andersen SAW. 2016. Notes from the field: secondary task precision for cognitive load estimation during virtual reality surgical simulation training. Eval Health Prof. 39:114–120.

Rop G, Wermeskerken M, Nooijer JA, Verkoeijen PPJL, Gog T. 2018. Task experience as a boundary condition for the negative effects of irrelevant information on learning. Educ Psychol Rev. 30:229–253.

Rotenstein LS, Ramos MA, Torre M, Segal JB, Peluso MJ, Guille C, Sen S, Mata DA. 2016. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. JAMA. 316:2214–2236.

Saleem JJ, Patterson ES, Militello L, Anders S, Falciglia M, Wissman JA, Roth EM, Asch SM. 2007. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. J Am Med Inform Assoc. 14:632–640.

Saus E, Johnsen BH, Eid J. 2010. Perceived learning outcome: the relationship between experience, realism and situation awareness during simulator training. Int Marit Health. 62:258–264.

Saus E, Johnsen BH, Eid J, Riisem PK, Andersen R, Thayer JF. 2006. The effect of brief situational awareness training in a police shooting simulator: an experimental study. Mil Psychol. 18:S3–S21.

Sawyer T, White M, Zaveri P, Chang T, Ades A, French H, Anderson J, Auerbach M, Johnston L, Kessler D. 2015. Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. Acad Med. 90:1025–1033.

Scerbo MW, Britt RC, Stefanidis D. 2017. Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. Am J Surg. 213:244–248.

Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. 2016. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. Med Educ. 50:682–692.

Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. 2017. Learner, patient, and supervisor features are associated with different types of cognitive load during procedural skills training: implications for teaching and instructional design. Acad Med. 92:1622–1631.

Sweller J, Ayres J, Kalyuga S. 2011. Cognitive load theory. New York (NY): Springer.

Sexton K, Johnson A, Gotsch A, Hussein AA, Cavuoto L, Guru KA. 2018. Anticipation, teamwork and cognitive load: chasing efficiency during robot-assisted surgery. BMJ Qual Saf. 27:148–154.

Shachak A, Hadas-Dayagi M, Ziv A, Reis S. 2009. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. J Gen Intern Med. 24:341–348.

Shewokis PA, Shariff FU, Liu Y, Ayaz H, Castellanos A, Lind DS. 2017. Acquisition, retention and transfer of simulated laparoscopic tasks using fNIR and a contextual interference paradigm. Am J Surg. 213:336–345.

Sibbald M, de Bruin AB, Cavalcanti RB, van Merrienboer JJG. 2013. Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. BMJ Qual Saf. 22:333–338.

Sibbald M, de Bruin ABH, van Merrienboer JJG. 2013. Checklists improve experts' diagnostic decisions. Med Educ. 47:301–308.

Siu KC, Suh IH, Mukherjee M, Oleynikov D, Stergiou N. 2010. The effect of music on robot-assisted laparoscopic surgical performance. Surg Innov. 17:306–311.

Stefanidis D, Wang F, Korndorffer JR, Jr., Dunne JB, Scott DJ. 2010. Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload. Surg Endosc. 24:377–382.

Sweller J. 1988. Cognitive load during problem solving: effects on learning. Cogn Sci. 12:257–285.

Szulewski A, Roth N, Howes D. 2015. The use of task-evoked pupillary response as an objective measure of cognitive load in novices and trained physicians: a new tool for the assessment of expertise. Acad Med. 90:981–987.

Thistlethwaite J, Hammick M. 2010. The Best Evidence Medical Education (BEME) Collaboration: into the next decade. Med Teach. 32:880–882.

Thomas L, Donohue-Porter P, Stein Fishbein J. 2017. Impact of interruptions, distractions, and cognitive load on procedure failures and medication administration errors. J Nurs Care Qual. 32:309–317.

Toff NJ. 2010. Human factors in anaesthesia: lessons from aviation. Br J Anaesth. 105:21–25.

Tomasko JM, Pauli EM, Kunselman AR, Haluck RS. 2012. Sleep deprivation increases cognitive workload during simulated surgical tasks. Am J Surg. 203:37–43.

Tremblay ML, Lafleur A, Leppink J, Dolmans DH. 2017. The simulated clinical environment: Cognitive and emotional impact among undergraduates. Med Teach. 39:181–187.

van Merrienboer JJ, Sweller J. 2010. Cognitive load theory in health professional education: design principles and strategies. Med Educ. 44:85–93.

Vandewaetere M, Manhaeve D, Aertgeerts B, Clarebout G, Van Merrienboer JJ, Roex A. 2015. 4C/ID in medical education: how to design an educational program based on whole-task learning: AMEE Guide No. 93. Med Teach. 37:4–20.

Walker J, von Bergmann H. 2015. Lessons from a pilot project in cognitive task analysis: the potential role of intermediates in preclinical teaching in dental education. J Dent Educ. 79:286–294.

Weigl M, Antoniadis S, Chiapponi C, Bruns C, Sevdalis N. 2015. The impact of intra-operative interruptions on surgeons' perceived workload: an observational study in elective general and orthopedic surgery. Surg Endosc. 29:145–153.

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. Surg Endosc. 30:559–566.

Weinger MB, Vredenburgh AG, Schumann CM, Macario A, Williams KJ, Kalsher MJ, Smith B, Truong PC, Kim A. 2000. Quantitative description of the workload associated with airway management procedures. J Clin Anesth. 12:273–282.

Workman M, Lesser MF, Kim J. 2007. An exploratory study of cognitive load in diagnosing patient conditions. Int J Qual Health Care. 19:127–133.

Wucherer P, Stefan P, Abhari K, Fallavollita P, Weigl M, Lazarovici M, Winkler A, Weidert S, Peters T, de Ribaupierre S, et al. 2015. Vertebroplasty performance on simulator for 19 surgeons using hierarchical task analysis. IEEE Trans Med Imaging. 34:1730–1737.

Young JQ, Boscardin CK, van Dijk SM, Abdullah R, Irby DM, Sewell JL, Ten Cate O, O'Sullivan PS. 2016. Performance of a cognitive load inventory during simulated handoffs: evidence for validity. SAGE Open Med. 4:2050312116682254.

Young JQ, Irby DM, Barilla-LaBarca M-L, ten Cate O, O'Sullivan PS. 2016. Measuring cognitive load: mixed results from a handover simulation for medical students. Perspect Med Educ. 5:24–32.

Young JQ, Niehaus B, Lieu SC, O'Sullivan PS. 2010. Improving resident education and patient safety: a method to balance initial caseloads at academic year-end transfer. Acad Med. 85:1418–1424.

Young JQ, Sewell JL. 2015. Applying cognitive load theory to medical education: construct and measurement challenges. Perspect Med Educ. 4:107–109.

Young JQ, Ten Cate O, O'Sullivan PS, Irby DM. 2016. Unpacking the complexity of patient handoffs through the lens of cognitive load theory. Teach Learn Med. 28:88–96.

Young JQ, Van Merrienboer J, Durning S, Ten Cate O. 2014. Cognitive load theory: implications for medical education: AMEE Guide No. 86. Med Teach. 36:371–384.

Young JQ, Wachter RM, Ten Cate O, O'Sullivan PS, Irby DM. 2016. Advancing the next generation of handover research and practice with cognitive load theory. BMJ Qual Saf. 25:66–70.

Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanström LL. 2010. Measuring mental workload during the performance of advanced laparoscopic tasks. Surg Endosc. 24:45–50.

Zheng B, Fu B, Al-Tayeb TA, Hao YF, Qayumi AK. 2014. Mastering instruments before operating on a patient: the role of simulation training in tool use skills. Surg Innov. 21:637–642.

Zheng B, Rieder E, Cassera MA, Martinec DV, Lee G, Panton ONM, Park A, Swanström LL. 2012. Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. Surg Endosc. 26:1352–1358.