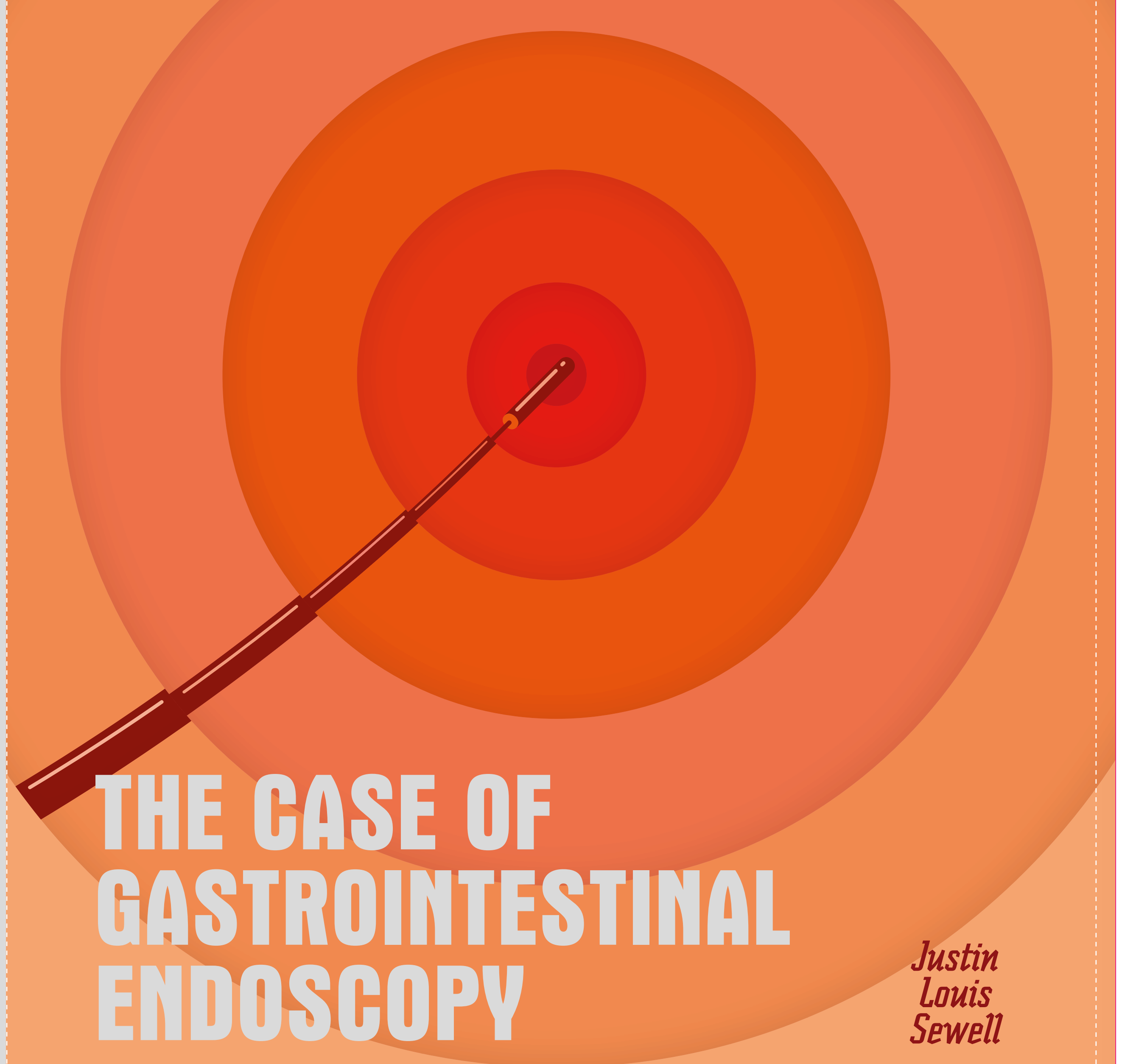
240 mm (final size)

246 mm (with bleed 3 mm)

COGNITIVE LOAD THEORY: A CRITICAL LENS FOR EXAMINING PROCEDURAL SKILLS TRAINING IN THE HEALTH PROFESSIONS THE CASE OF GASTROINTESTINAL ENDOSCOPY Justin Louis Sewell

COGNITIVE LOAD THEORY:

A CRITICAL LENS FOR EXAMINING PROCEDURAL SKILLS TRAINING IN THE HEALTH PROFESSIONS



THE CASE OF GASTROINTESTINAL ENDOSCOPY

Justin Louis Sewell

Invitation

to attend the public defense of my thesis

COGNITIVE LOAD THEORY:

A CRITICAL LENS FOR EXAMINING PROCEDURAL SKILLS TRAINING IN THE HEALTH PROFESSIONS

THE CASE OF GASTROINTESTINAL ENDOSCOPY

30 August 2019 | 2:30 PM

Senate Hall, Academy Building
Domplein 29, Utrecht
Utrecht University
Reception in the Academy Building immediately following the defense

Justin Louis Sewell
justin.sewell@ucsf.edu

230 mm (final size)

236 mm (with bleed 3 mm)

170 mm (final size)

173 mm (with bleed 3 mm)

15 mm

170 mm (final size)

173 mm (with bleed 3 mm)

65 mm

71 mm

Cognitive Load Theory: A Critical Lens for Examining Procedural Skills Training in the Health Professions

The Case of Gastrointestinal Endoscopy

Justin Louis Sewell

ISBN: 978-94-6375-352-4

Cover art courtesy of Mark van Wijk.

Cover: Ridderprint BV, the Netherlands
Printing: Ridderprint BV, the Netherlands

Cognitive Load Theory: A Critical Lens for Examining Procedural Skills Training in the Health Professions

The Case of Gastrointestinal Endoscopy

Cognitive Load Theory: een kritisch perspectief voor onderzoek naar
de training in procedurele vaardigheden bij beroepen in de gezondheidszorg,
toegepast op de gastro-intestinale endoscopie

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op vrijdag 30 augustus 2019 des middags te 2.30 uur

door

Justin Louis Sewell

geboren op 14 oktober 1976 te Abilene, Texas, Verenigde Staten van Amerika

Promotoren:

Prof. dr. Th.J. ten Cate

Prof. dr. P.S. O'Sullivan

Copromotor:

Dr. C.K. Boscardin

Table of Contents

PREFACE

I started my gastroenterology fellowship training immediately after completing a three-year residency in a demanding Internal Medicine program that afforded residents a great deal of autonomy. I considered myself to be a talented clinician, having managed the care of hundreds of patients, many critically ill. And yet, when I was introduced to the practice of gastrointestinal (GI) endoscopy, I felt that I had returned to the drawing board. Despite a brain brimming with knowledge about the human body, I had no framework through which to conceptualize what was happening inside the bodies of my patients when I inserted an endoscopic instrument. My teachers offered advice intended to be helpful, such as *"Your scope is too long,"* or, *"You have a loop. You need to reduce the loop and get the scope nice and straight. Feel the tension go out as you reduce."* While I understood the individual words they were saying, I did not understand what they actually meant, or how to apply their teaching.

Of course, I did eventually learn what they meant, as all successful procedural learners do, but it took a great deal of time and involved a lot of frustration and heartache. I developed adequate endoscopic competence and graduated from fellowship. I walked out of the door on June 30 as a learner and returned on July 1 as a teacher! Not surprisingly, I found myself using similar words and similar teaching techniques, and, also not surprisingly, I saw on my new learners' faces similar glazed expressions and blank stares. I witnessed similar heartache and frustration among my learners that I had experienced so recently. "*Why does this have to be so difficult?"* I wondered.

These challenges as both learner and teacher of medical procedures directly inspired me to pursue training in teaching techniques and education research through the Teaching Scholars Program at the University of California (UCSF). During this program, I discovered cognitive load theory (CLT), a cognitive learning theory focused on the limitations of working memory. CLT seemed highly authentic to my experience as a learner and teacher of procedural skills, and I wanted to study its role in GI endoscopy training. Although Teaching Scholars introduced me to education research, I knew that I needed more intensive training to achieve my goals, and so I matriculated to the UMC Utrecht-UCSF Doctoral Program in Health Professions Education. The fruits of my training and research endeavors are detailed in this thesis, which investigates teaching and learning of procedural skills in the health professions through the lens of CLT. The following Introduction sections will introduce the reader to the framework of CLT, the challenges of procedural skills teaching and learning, and how CLT applies to this unique area of health professions education (HPE).

Chapter 1

Introduction

COGNITIVE LOAD THEORY

CLT is a cognitive learning theory based on the Atkinson-Shriffin model of memory[1] (Figure 1). First described in 1968, this model proposes three types of human memory. Sensory memory includes all the stimuli that human senses can detect. In classroom learning settings, vision and hearing are the most commonly used, but in the health professions, touch, and even smell, can be important as well.

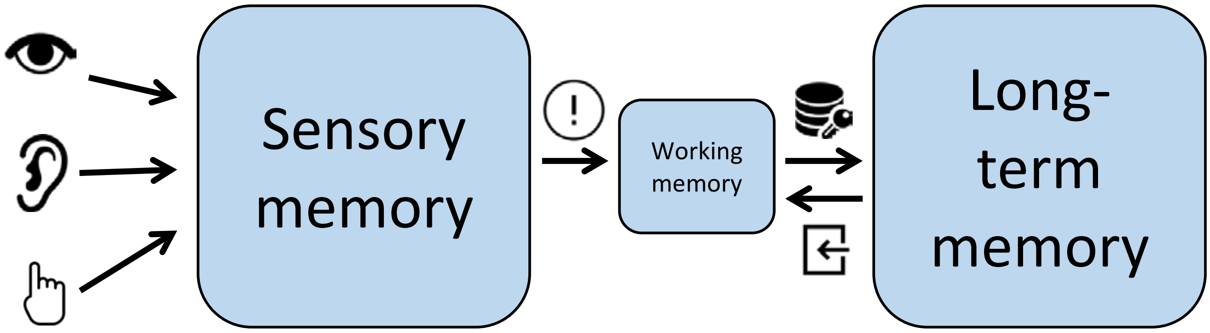


Figure 1. Model of human memory. Adapted from Young JQ et al, Medical Teacher 2014;36(5):371-384. Permission obtained from Taylor & Francis Group, http://www.tandfonline.com.

Because humans are constantly bombarded with sensory stimuli, we must use working memory to select which stimuli to direct our attention. In the education setting, attending to the wrong stimuli will interfere with learning. Some information is transferred from working memory to long-term memory in the form of learning schema (i.e., organized bundles of information that together explain a concept or construct), which can later be retrieved for use. Of vital importance is the significant limitation of working memory. While capacity for both sensory and long-term memory is considered infinite, working memory is extremely limited in both capacity and duration, and can manage only a few pieces of information at any given time. Prior work has suggested working memory can manage around 7 +/- 2 informational elements at any given time[2], though more contemporary research suggests the capacity is probably smaller, especially when the information involves a high degree of element interactivity or when active processing of information is required[3]. Furthermore, information in working memory is rapidly lost if it is not practiced and transferred to long-term memory as a schema. Working memory is generally considered to be largely fixed in capacity and has even been proposed by some as a primary contributor to the construct of intelligence[4]. Thus, educators must ensure that their teaching does not overload the limited working memory; if this occurs, both learning and performance will suffer[5].

CLT was first described by Sweller in 1988[6] and subsequently elaborated on by others[5,7-9]. CLT builds on the Atkinson-Shriffin model by proposing three types of mental activities that occupy the working memory. These three types of cognitive load, intrinsic load, extraneous load and germane load, all contribute to working memory demands. Working memory bandwidth utilized to directly complete essential elements of a learning task is termed *intrinsic load*. Intrinsic load will be higher when learners have less experience, or when tasks are more complex. If intrinsic load is too high, the learner's working memory will become overwhelmed, whereas if the intrinsic load is too low, learning will be limited as there is nothing for the learner to learn. Therefore, intrinsic load should be matched to the learner's experience level and zone of proximal development[5,10]. When learners pay attention to, or focus on, thoughts or stimuli not essential to the learning task, another type of cognitive load, *extraneous load*, occurs. Distractions induce extraneous load, and can be external within the learning environment (e.g., conversations not related to the learning task, mobile devices vibrating or ringing, other interruptions) or internal within the learner's mind (e.g., negative emotions, internal distractions)[5,11,12]. Such distractions occupy space in working memory while they occur and for a period of time afterward. Teachers may inadvertently induce extraneous load. Classic examples are mismatches in information between sensory channels (e.g., a lecturer speaks about a topic not matching her slides) or the use of unfamiliar jargon or terminology. In either of these cases, the learner must focus working memory resources on the modality or delivery of teaching, rather on the content of teaching, which reduces capacity for completing and learning from the task. Because extraneous load always diminishes space for productive working memory activities, it should be minimized. Simply completing a task does not automatically lead to learning; for learning to occur, learners must engage in the third type of cognitive load, which is *germane load*. Germane load occurs when learners use working memory resources to form or refine learning schema, including connecting new knowledge to existing schema. Germane load should be optimized because it leads to learning. In order for this to occur, the other two goals of CLT must be met: to match intrinsic load to the learner's level, and to minimize extraneous load. In the motivated learner, this provides space for germane load and learning.

Working memory limitations constrain the amount of the three cognitive load types. For example, if intrinsic and/or extraneous overload working memory (as in Figure 2, part A), there is no spaced for germane load. By adjusting the amount of extraneous or intrinsic load (as in Figure 2, part B), space is opened for germane load and learning. A variety of teaching strategies are thought to promote germane load, such as asking clarifying questions, reflection and task variability[5].

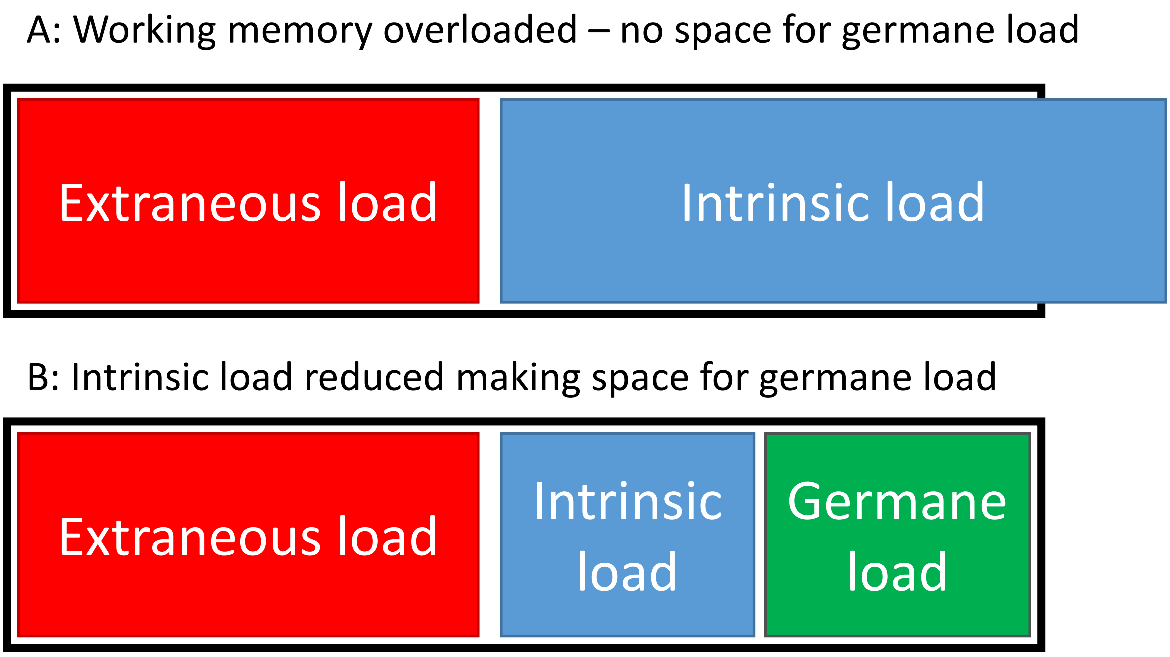


Figure 2. Cognitive load types and working memory constraints. Adapted from Young JQ et al, Medical Teacher 2014;36(5):371-384. Permission obtained from Taylor & Francis Group, http://www.tandfonline.com.

Since its original description, many scholars have studied CLT in experimental research involving non-HPE classroom settings[13]. Given its focus on working memory limitations, CLT is well suited for complex learning settings, which are commonly found in HPE. Furthermore, its attention to tasks, environments, and learners provides a framework relevant to HPE learning. Consequently, HPE researchers are increasingly interested in applying CLT to HPE learning. Given the relative youth of CLT as a theory, and its fairly recent application to HPE, its role in HPE learning is still being characterized. Application of CLT to controllable learning environments like simulation[14,15] is better-studied than less controllable workplace environments. Indeed, as Chapter 4 will demonstrate, study of CLT in HPE workplace settings has been quite narrowly focused, with few examples of how to use CLT to improve HPE workplace learning. The work in this thesis is intended to advance our understanding of how CLT may be applied within complex HPE workplace settings, particularly procedural settings.

Currently, there are two noteworthy debates among CLT scholars. The first relates to the three types of cognitive load; these remain theoretical, as it is difficult to empirically prove what exactly occupies a learner's working memory at any moment in time. While the CLT community agrees upon intrinsic and extraneous load as incontrovertible types of cognitive load, whether germane load is a unique construct, or whether it should be conceptualized as working memory resources used to manage intrinsic load, remains the subject of debate[16,17]. This has been a difficult debate to address; as Chapter 4 will demonstrate, germane load remains understudied, at least in workplace settings. The second primary debate is how and when to measure or estimate cognitive load. The three primary modalities are psychometric self-rating instruments (either generic or task-specific; most commonly used being Paas' single item[7] and the NASA Task Load Index[18]), response to a secondary task (which will be slower and less accurate when cognitive load is higher), and physiological measurements (such as heart rate variability, skin conductance, eye movement, electroencephalogram, or functional MRI)[5]. Each of these measurement approaches has benefits and drawbacks in terms of cost, disruption to the learning environment, and type of data provided. Currently, the only technique that can assess different types of cognitive load (as opposed to overall cognitive load or mental effort), is the use of psychometric self-rating instruments. The research described in this thesis informs both of the debates described above, and the debates will be addressed in the Discussion section.

PROCEDURAL SKILLS IN THE HEALTH PROFESSIONS

This thesis will consider how CLT informs procedural skills teaching and learning in the health professions. Although the idea of "cognitive procedures" (e.g., developing differential diagnosis, clinical reasoning) has been described[19], we focus on psychomotor procedures, that is, procedures requiring learners to use parts of their body (usually hands but sometimes other body parts as well) to accomplish a physical task which in some way alters or invades the patient's body in order to diagnose, treat, or prevent disease. (Based on this definition, we exclude physical examination tasks which require physical contact with the patient but do not invade or alter the patient's body; for example, cardiac auscultation or examination of the knee.) The spectrum of procedural tasks performed across the health professions is vast, ranging in duration, complexity, and risk. At one end of the spectrum is a simple task like venipuncture, which takes seconds, is very low risk, and can be accomplished by a provider after very limited training. At the other end is a complex multistage operation that takes hours, requires multiple ultra-specialized surgeons, and has life-or-death stakes. Between these extremes exists a vast array of thousands of different procedures, all of which have the potential to promote or improve health, but also have risk of harm to patients.

Though there is great heterogeneity in procedural skills, they are linked together as a group by the perspectives of three major stakeholders: patients, learners, and teachers. The challenges inherent to each of these stakeholder groups sets procedural skills apart from other HPE learning. Indeed, the teaching and learning of procedures is among the most challenging activities with in HPE[20]!

For patients, the primary issues are discomfort, embarrassment, and risk. When a patient allows a health professional to perform a procedure, she trusts the proceduralist with her body, because the proceduralist is literally invading her body's defense mechanisms. For many procedures, the patient will be partially or completely sedated, and so she will be unaware of what is happening, increasing the need for trust in the provider. Furthermore, many procedures (such as those studied in this thesis) involve patient's most private body parts, and so the procedures may be a source of embarrassment.

For learners, the most evident unique challenges are the psychomotor components of a task; the learner must learn to skillfully use of her hands (and other body parts as well for some tasks) to help and not harm the patient. Many psychomotor tasks require extreme accuracy, and many learners may have limited prior experience using their hands in such coordinated fashion; in other words, the essential task demands (intrinsic load) are high. Furthermore, learners may be unfamiliar with psychomotor terminology (a source of extraneous load) and may struggle to follow their teachers' directions. While learning *how* to perform the psychomotor components of the task, the learner must also learn *when* and *why*. That is, many procedural skills entail not only significant psychomotor components, but also significant cognitive components. This presents a "dual-task" condition, which increases intrinsic cognitive load[13,21]. Cognitive decisions that must be made while performing procedures may in themselves be quite complex, and sometimes must be made within high-pressure environments. During a procedure, the providers rarely have the luxury of time to think extensively about a decision. Whether it is an emergency procedure, or simply a routine procedure in the midst of a busy operating room schedule, decisions made during procedures generally have to be made quickly, yet at the same time there are high stakes, because the wrong decision could render harm, either immediately or in the future. Time pressure is known to decrease accuracy of decision-making and likely contributes to extraneous load [22,23]. The chaotic or pressured environment of some HPE procedures (e.g., operating room, emergency department) may further contribute to extraneous load. Furthermore, for many learners, the experience of learning procedural skills may be highly frustrating. Most HPE learners have spent years honing cognitive skills, and so learning cognitive skills is likely more familiar, even when the content is complex and challenging. Conversely, their lack of experience with psychomotor skills may lead to frustration when learning procedural skills, some of which take several hundred repetitions before mastery is possible[24]. Simply performing a procedure may be so mentally overwhelming that the learner has no working memory space available to hear his teacher's instructions, much less to actually learn from the experience. Finally, as if the above were not challenging enough, modern HPE learners have numerous competing demands that have potential to induce extraneous load and interfere with learning.

Teachers may also experience substantial challenges as they teach procedural skills in the health professions, because they must balance the healthcare needs of the patient, the learning needs of the learner, the efficiency and quality demands of the healthcare system, and their own internal desires and level of patience. On a daily basis, procedural teachers are faced with difficult decisions about when and how and with what tasks to entrust procedural learners[25]. In addition, teachers must learn to articulate complex psychomotor concepts and decide what portion(s) of a procedure for which a learner has adequate experience to perform. Essentially, a procedural skills teacher must skillfully orchestrate a complex interaction with multiple stakeholders to accomplish a technically successful outcome while teaching one or more learners and following time constraints!

MODELS OF PROCEDURAL SKILLS LEARNING

No single theory or model of motor skill learning has been found as fully representing or explaining the complexities of motor skill development[26], yet certain learning paradigms have been studied in relation to procedural skills learning.

While Skinner's behaviorist approach[27] has long been a theoretical underpinning of procedural teaching, most modern research has turned towards more cognitive approaches. Modern evidence suggests that Halsted's classic approach of "see one, do one, teach one" is likely inapplicable to modern HPE[28], yet other models have been proposed. Fitts and Posner proposed that procedural learners progress through three stages of psychomotor learning: cognitive, integrative/associative, and autonomous[29]. In the cognitive phase, learners learn *about* the procedure and may observe others performing the procedure and make basic attempts toward first psychomotor tasks. In the integrative or associative phase, learners begin to refine accuracy and efficiency of movements. In the autonomous phase, movements are accurate and efficient. This model has been applied to a wide variety of psychomotor learning settings, ranging from track and field events[30] to HPE (including surgery[31] and GI endoscopy[32]). While Fitts and Posner focused on progress of the learner, Peyton suggested four instructional steps to promote procedural learning: demonstration, deconstruction, comprehension, performance[33]; application of Peyton's model has been shown to benefit learning in medical procedures[28].

Ericsson's theory of deliberate practice for mastery learning has also been studied extensively as relating to motor learning[34]. Ericsson's theory includes ten tenets of deliberate practice and eight properties of mastery learning that can be applied to promote learning of motor skills. A systematic review suggested that applying Ericsson's model may promote improvements in actual patient care processes and outcomes[35], yet it is more focused at the meta-level of curricular design and its complexity (ten features of deliberate practice, eight properties of mastery learning[36]) than direct procedural teaching, and may be challenging for application among everyday clinical educators lacking formal education training.

COGNITIVE LOAD THEORY AND PROCEDURAL SKILLS TRAINING

The above models clearly inform crucial elements of procedural skills learning, yet they address learning and skill acquisition at a relatively meta level. Conversely, CLT focuses directly on the learner, at the most central and limited element of learning (from the cognitive perspective): the working memory. Therefore, CLT provides a perspective complementary to those described above. We consider CLT as highly relevant to procedural skills training for several reasons. CLT is considered most applicable to complex learning settings[5]. Considering the dual task setting and numerous demands imposed on learners as described above, procedural training settings may well be among the most complex in all of HPE; in many procedural settings, there is tremendous potential for cognitive overload. Intrinsic load may be substantially elevated due to complexity of cognitive and motor task elements. Extraneous load may be high due to a chaotic procedural environment, urgent or emergent procedural setting, learner stress over their own performance and achievement, or confusing terminology as teachers try to articulate complex motor concepts in words that learners can understand. If learners are routinely overloaded by the task complexity and by distractions and unproductive cognitive processing, they are being set up to have minimal working memory capacity for germane load and learning. As complexities of medical procedures advances and duty hours and other training restrictions clamp down, it is vital that procedural learning in the health professions is optimized. CLT provides a useful lens through which to study possibilities to optimize capacity and space for learning procedural skills. Given its specific focus on the learner's working memory, CLT may complement other learning models and paradigms such as those of Fitts and Posner, and Ericsson, and be used to promote learning.

GASTROINTESTINAL ENDOSCOPY AS AN EXEMPLAR SETTING

As discussed above, there exists substantial heterogeneity in the types of procedural skills learned and performed in the health professions. To better understand the role of CLT in procedural skills learning, we therefore sought to select an exemplar setting, ideally a "middle-of-the-road" class of procedures that was neither too simple nor too complex, neither too short nor too long. We also sought to select a class of procedures with substantial impact on the health of populations. We identified GI endoscopy as an ideal exemplar procedural setting in which to study the role of CLT, as described below. While a few scholars have linked CLT to GI endoscopy training[32,37,38], empirical evidence for CLT in GI endoscopy training has been very limited.

GI endoscopy encompasses a set of moderately invasive procedures, usually performed under sedation, through insertion of video-equipped steerable tubular instruments ("endoscopes") into the mouth or anus of a patient (Figures 3a and 3b).

Figure 3a. A colonoscope.

Figure 3b. A selection of gastrointestinal endoscopes.

This equipment allows the performing clinician to examine multiple organs including the esophagus, stomach, small intestine, bile ducts, pancreas ducts, and colon. Essentially the entire GI tract can be accessed via endoscopic procedures (although some parts are more challenging to reach than others). The most common GI endoscopic procedures are esophagogastroduodenoscopy, which examines the upper GI tract, and colonoscopy (Figure 4), which examines the lower GI tract and can substantially reduce morbidity and mortality from colorectal cancer[39]. Several other types of GI endoscopic procedures exist as well, including highly specialized procedures that can examine and treat diseases of the biliary tract and pancreas (e.g., endoscopic retrograde cholangiopancreatography and endoscopic ultrasound). GI endoscopy is most commonly performed by gastroenterologists, but some surgeons and general practitioners also perform GI endoscopies. Although physicians most commonly perform GI endoscopic procedures, mid-level providers such as nurse practitioners do as well[40].

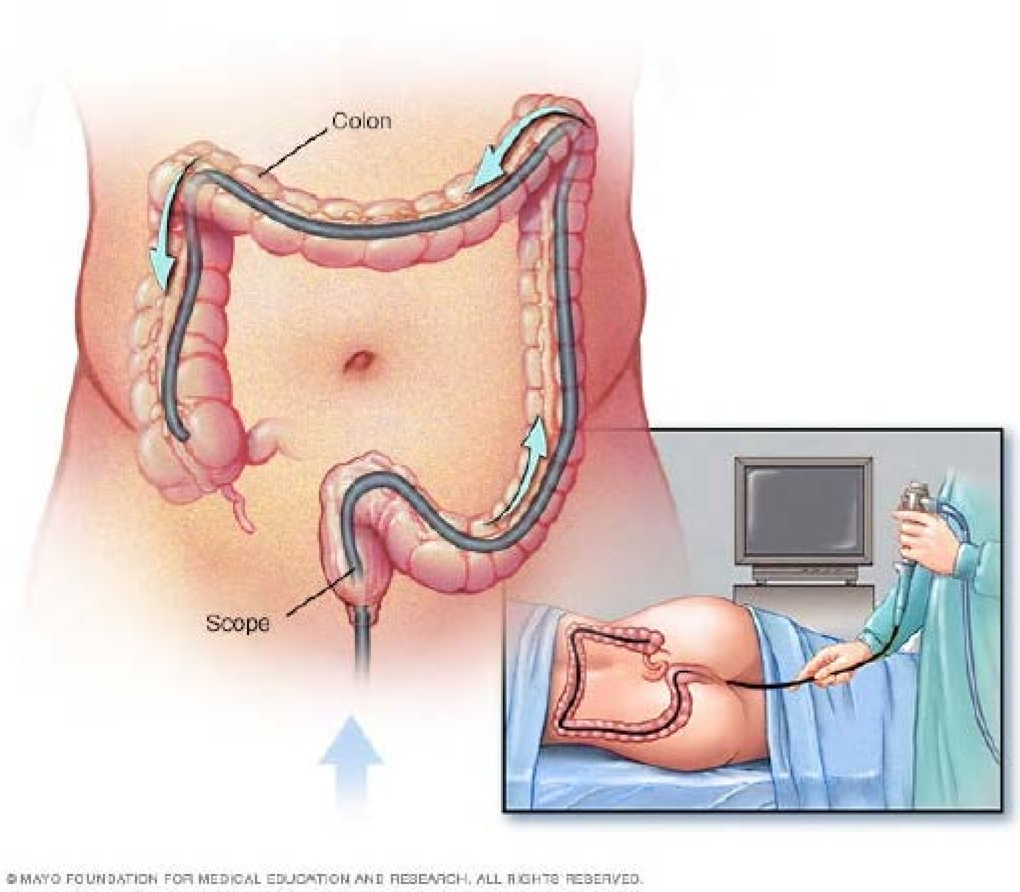


Figure 4. Colonoscopy technique. Used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

GI endoscopic procedures involve all of the general procedural learning challenges described previously and have the additional challenges of substantial limitations in the visual and tactile channels. The endoscopic instrument assumes an ever-changing three-dimensional configuration within the patient. This configuration can be complex, and can form “loops” (Figure 5, part B). Small loops can cause patients to have pain, while large loops can induce serious complications such as bleeding, perforation or rupture of the intestine. However, the endoscopist’s visual channel is limited to the two-dimensional, cross-sectional video transmitted by the instrument tip[41]. The endoscopist's tactile information is constrained as well. Endoscopes are long and flexible, and therefore provide much less direct haptic feedback compared with direct hands-on procedures, or procedures with short, rigid instruments (such as laparoscopy)[41]. The endoscopist must therefore use limited visual and tactile input to intuit the complex configuration of the long endoscopic instrument. These constraints in sensory feedback significantly increase challenges of learning GI endoscopy (i.e., intrinsic load is high), reflected by the fact that around 300 procedures must be completed before competence can be attained[42]. In addition to significant motor demands, many endoscopic procedures have substantial cognitive demands, and some are performed in emergent or otherwise pressurized or distracting settings. Essentially, GI endoscopy training is a setup for cognitive overload, yet how we should best train future endoscopists has not been well characterized.

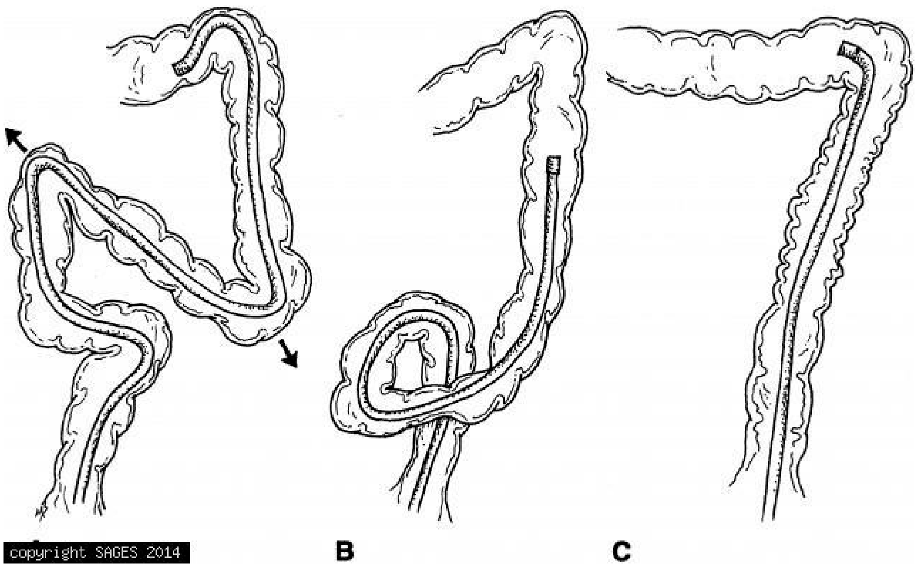


Figure 5. Colonoscopy loops. Used with permission from Society of American Gastrointestinal and Endoscopic Surgeons (SAGES). https://www.sages.org/image-library/formation-of-loops-in-the-colon/.

We consider GI endoscopy as residing near the middle of the spectrum of complexity for HPE procedures. It is more involved and complex than simpler bedside procedures often performed by nurses (e.g., peripheral IV insertion, bladder catheter insertion) or general internists (e.g., central venous catheter insertion, paracentesis) yet less complex and more readily mastered than many surgical operations. Indeed, GI endoscopy occupies the crossroads of surgical and non-surgical procedures, and this "middle-of-the-road" status makes it an ideal setting in which study of CLT may produce findings that are generalizable to both more and less complex procedural settings.

Aspects of how GI endoscopy is taught and practiced are also advantageous for education research. Teaching of GI endoscopy is generally hands-on (often starting with simulation followed by practice on real patients) in a 1:1 teacher: learner setting. This facilitates direct analysis of how teaching practices may impact cognitive load and learning. GI endoscopic procedures are performed very commonly, providing ample examples to study; at the author's home program, each gastroenterology fellow is expected to perform more than 300 procedures in their first year alone. Additionally, GI endoscopy represents a defined task with a clear beginning, end, and intermediate steps. The vast majority of GI procedures take well under 1 hour to complete, making study more feasible than for longer procedures. The goals of each GI endoscopy are clear, and established performance metrics with validity evidence exist[43].

Although the rate of HPE research publication continues to accelerate[44], study of how trainees learn to perform GI endoscopy remains relatively stagnant, focused largely on learning curves[45], assessment of performance and competence[24,43,46], impact of trainee involvement on clinical outcomes[47], and how simulation impacts endoscopy learning[48]. We find that modern endoscopy training remains largely eminence-based, rather than evidence-based, relying largely on a traditional apprenticeship model of learning. The work in this thesis is meant to advance the science of how to teach both GI endoscopy, and procedural skills in general.

THESIS OUTLINE

Considering the complexities of procedural skills training and potential for cognitive overload, CLT presents an ideal theoretical lens through which to study procedural skills training, yet we appreciate that rigorous study of CLT in procedural settings has been limited. Major gaps include how to measure intrinsic, extraneous and germane load in procedural settings, what factors contribute to these cognitive load subtypes in procedural settings, and how teaching can be leveraged to optimize cognitive load among procedural learners.

The research in this thesis follows a logical progression that addresses these gaps. We first document efforts to develop an instrument to measure intrinsic, extraneous and germane load during procedural training and then consider factors associated with each cognitive load subtype. Having established measurement methods and some understanding of contributors to cognitive load in procedural training settings, we next turn our attention to teaching approaches, first by consulting the existing literature, and then by performing qualitative and mixed methods research with both procedural teachers and learners. We focus on GI endoscopy as an exemplar setting, which serves as a point of focus to address deep questions related to CLT and procedural skills training. Resurfacing from that deep dive will enable extrapolation and generalization to diverse procedural settings.

In order to accomplish our goals, we performed five studies addressing the following five research questions:

1) How can we measure intrinsic, extraneous and germane cognitive load among colonoscopy learners?
2) How do features of learners, tasks, settings and teachers influence intrinsic, extraneous and germane cognitive load during colonoscopy training?
3) How has CLT been used to study teaching and learning within professional workplace settings, both within and outside of the health professions?
4) What learning challenges do experienced GI endoscopy teachers see their learners experience, what teaching strategies do they use to help learners overcome these challenges, and how do those strategies align with cognitive load theory?
5) How do "everyday" teachers of GI endoscopy teach, and how does their teaching affect learners' perceived intrinsic, extraneous and germane cognitive load?

We addressed these research questions as follows. To study cognitive load during GI endoscopy training, we must be able to measure cognitive load. Although simple scales to estimate overall cognitive load exist[7], we consider the ability to separately assess intrinsic, extraneous and germane cognitive load to be of great importance for furtherance of CLT and to develop practical recommendations. Therefore, the first study (Chapter 2) was designed to rigorously develop and collect evidence for validity of a self-report psychometric instrument (the Cognitive Load Inventory for Colonoscopy, or CLIC) to estimate the three different types of cognitive load among gastroenterology fellows learning to perform colonoscopy. In the second study, we sought to identify characteristics of learners, tasks, settings and teachers that might influence the three cognitive load types among colonoscopy learners (Chapter 3); we designed a study to model unique contributors to each cognitive load type. Having established relevance of CLT to colonoscopy training and armed with a better understanding of contributors in the GI endoscopy setting, we next turned our attention to how teaching could affect cognitive load in procedural settings. To leverage the existing literature, we designed a comprehensive Best Evidence in Medical Education (BEME) Review of studies using CLT to study workplace teaching among professions within, and outside, the health professions (Chapter 4). We next sought to understand teaching approaches used during GI endoscopy training by performing a qualitative study interviewing experienced endoscopy educators within the United States, Canada and the Netherlands. We asked them about challenges they saw their GI endoscopy learners experience, and teaching strategies they used to help learners overcome these challenge (Chapter 5); we used CLT as the primary theoretical lens for qualitative analysis. For the fifth and final study, we sought to understand how procedural teaching strategies were enacted among "everyday" teachers of GI endoscopy at two large hospitals, and how they affected learner perceptions of intrinsic, germane and extraneous cognitive load (Chapter 6). We observed teaching that occurred between attending physicians (teachers) and gastroenterology fellows (learners), used real-time content analysis during the colonoscopy to identify the teaching activities that were used, and interviewed fellows regarding perceived impact on cognitive load. The study also included a single provocative quantitative analysis linking instances of germane load-promoting teaching activities with measured germane load. The thesis concludes with a discussion of how the forgoing chapters advance our understanding of procedural skills training in the health professions and cognitive load theory, and how CLT can inform procedural skills training (Chapter 7); the generalizability of GI endoscopy to other procedural skills is emphasized. The discussion chapter specifically addresses the two debates related to CLT noted above: how to measure cognitive load and the “true identity” of germane load.

Note: This thesis is a collection of related research articles, with each chapter being written as a stand-alone article. Therefore, some degree of repetition and overlap across chapters is expected.

REFERENCES

1. Atkinson RCS, R.M. Human memory: A proposed system and its control processes. In: Spence KWTS, J., ed. *Psychology of Learning and Motivation.* New York: Academic Press; 1968:89-195.
2. Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev.* 1956;63(2):81-97.
3. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci.* 2001;24(1):87-114; discussion 114-185.
4. Rosman T, Mayer A-K, Krampen G. A longitudinal study on information-seeking knowledge in psychology undergraduates: Exploring the role of information literacy instruction and working memory capacity. *Comput Educ.* 2016;96:94-108.
5. Young JQ, van Merriënboer J, Durning S, ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Med Teach.* 2014;36(5):371-384.
6. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci.* 1988;12:257-285.
7. Paas F, Tuovinen JE, Tabbers H, van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol.* 2003;38(1):63-71.
8. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010;44(1):85-93.
9. van Merriënboer JJG, Kirschner PA, Kester L. Taking the load off a learner's mind: Instructional design for complex learning. *Educ Psychol.* 2003;38(1):5-13.
10. Kostons D, van Gog T, Paas F. Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learn Instr.* 2012;22(2):121-132.
11. Fraser K, Huffman J, Ma I, et al. The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. *Chest.* 2014;145(5):958-963.
12. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ.* 2012;46(11):1055-1062.
13. Feldon DF. Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educ Psychol.* 2007;42(3):123-137.
14. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ.* 2015;48(8):805-814.

15. Naismith LM, Cavalcanti RB. Validity of Cognitive Load Measures in Simulation-Based Training: A Systematic Review. *Acad Med.* 2015;90(11 Suppl):S24-35.
16. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. *Perspect Med Educ.* 2015;4(3):107-109.
17. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. *Perspect Med Educ.* 2015;4(3):119-127.
18. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc Hum Factors Ergon Soc Annu Meet.* 2006;50(9):904-908.
19. Beaunieux H, Hubert V, Witkowski T, et al. Which processes are involved in cognitive procedural learning? *Memory.* 2006;14(5):521-539.
20. Grantcharov TP, Reznick RK. Teaching procedural skills. *BMJ.* 2008;336(7653):1129-1131.
21. Mayer RE, Moreno R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educ Psychol.* 2003;38(1):43-52.
22. Modi H, Leff D, Singh H, Darzi A. The influence of cognitive load on technical ability among surgical trainees. *Int J Surg.* 2016;36:S34-S35.
23. Kataoka J, Sasaki M, Kanda K. Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Jpn J Nurs Sci.* 2011;8(1):47-56.
24. Sedlack RE, Coyle WJ. Assessment of competency in endoscopy: establishing and validating generalizable competency benchmarks for colonoscopy. *Gastrointest Endosc.* 2016;83(3):516-523.e511.
25. ten Cate O. Supervision and Entrustment in Clinical Training: Protecting Patients, Protecting Trainees. *AHRQ Web M&M.* 2018. Available online at https://psnet.ahrq.gov/webmm/case/461/Supervision-and-Entrustment-in-Clinical-Training-Protecting-Patients-Protecting-Trainees. Accessed 29 January 2019.
26. Newell KM. Motor skill acquisition. *Annu Rev Psychol.* 1991;42:213-237.
27. Skinner BF. Operant behavior. *Am Psychol.* 1963;18(8):503-515.
28. Romero P, Gunther P, Kowalewski KF, et al. Halsted's "See One, Do One, and Teach One" versus Peyton's Four-Step Approach: A Randomized Trial for Training of Laparoscopic Suturing and Knot Tying. *J Surg Educ.* 2018;75(2):510-515.
29. Fitts PM, Posner MI. *Human Performance.* Oxford: Brooks & Cole; 1967.
30. Taylor JA, Ivry RB. The role of strategies in motor learning. *Ann N Y Acad Sci.* 2;1251:1-12.
31. Bugdadi A, Sawaya R, Olwi D, et al. Automaticity of Force Application

During Simulated Brain Tumor Resection: Testing the Fitts and Posner Model. *J Surg Educ.* 2018;75(1):104-115.
32. Coderre S, Anderson J, Rostom A, McLaughlin K. Training the endoscopy trainer: from general principles to specific concepts. *Can J Gastroenterol.* 2010;24(12):700-704.
33. Gradl-Dietsch G, Lubke C, Horst K, et al. Peyton's four-step approach for teaching complex spinal manipulation techniques - a prospective randomized trial. *BMC Med Educ.* 2016;16(1):284.
34. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004;79(10 Suppl):S70-81.
35. Griswold-Theodorson S, Ponnuru S, Dong C, Szyld D, Reed T, McGaghie WC. Beyond the simulation laboratory: a realist synthesis review of clinical outcomes of simulation-based mastery learning. *Acad Med.* 2015;90(11):1553-1560.
36. McGaghie WC, Kristopaitis T. Deliberate practice and mastery learning: origins of expert medical performance. In: Cleland J, Durning SJ, eds. *Researching Medical Education.* Vol I. West Sussex, UK: John Wiley & Sons, Ltd; 2015:219-230.
37. Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. 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.* 2014;28(3):155-159.
38. Coderre S, Anderson J, Rikers R, Dunckley P, Holbrook K, McLaughlin K. Early use of magnetic endoscopic imaging by novice colonoscopists: Improved performance without increase in workload. *Can J Gastroenterol.* 2010;24(12):727-732.
39. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L. Association of colonoscopy and death from colorectal cancer. *Ann Intern med.* 2009;150(1):1-8.
40. Baumgardner JM, Sewell JL, Day LW. Assessment of quality indicators among nurse practitioners performing upper endoscopy. *Endosc Int Open.* 2017;5(9):E818-e824.
41. Kaltenbach T, Leung C, Wu K, Yan K, Friedland S, Soetikno R. Use of the colonoscope training model with the colonoscope 3D imaging probe improved trainee colonoscopy performance: a pilot study. *Dig Dis Sci.* 2011;56(5):1496-1502.
42. Spier BJ, Benson M, Pfau PR, Nelligan G, Lucey MR, Gaumnitz EA. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc.* 2010;71(2):319-324.

43. ASGE Training Committee, Sedlack RE, Coyle WJ, et al. ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD. *Gastrointest Endosc.* 2014;79(1):1-7.
44. Ji YA, Nam SJ, Kim HG, Lee J, Lee SK. Research topics and trends in medical education by social network analysis. *BMC Med Educ.* 2018;18(1):222.
45. Siau K, Kuwai T, Ishaq S. Analysis of learning curves in gastroscopy training: the need for composite measures for defining competence. *Gut.* 2018;67(6):1198.
46. Kumar NL, Kugener G, Perencevich ML, Saltzman JR. The SAFE-T assessment tool: derivation and validation of a web-based application for point-of-care evaluation of gastroenterology fellow performance in colonoscopy. *Gastrointest Endosc.* 2018;87(1):262-269.
47. Rustagi T. The influence of fellow involvement and level of fellowship training on adenoma detection rates. *Gastrointest Endosc.* 2010;72(5):1114-1115; author reply 1115.
48. Singh S, Sedlack RE, Cook DA. Effects of Simulation-Based Training in Gastrointestinal Endoscopy: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol.* 2014;12(10):1611-1623.e1614.

Chapter 2

Measuring cognitive load during procedural skills training with colonoscopy as an exemplar

Sewell, JL

Boscardin, CK

Young, JQ

ten Cate, O

O'Sullivan, PS

Published as:

Measuring cognitive load during procedural skills training with colonoscopy as an exemplar

Medical Education 2016;50(6):682-692

ABSTRACT

Objectives: Few studies have investigated cognitive factors affecting learning of procedural skills in medical education. Cognitive load theory, which focuses on working memory, is highly relevant, but methods for measuring cognitive load during procedural training are not well-understood. Using colonoscopy as an exemplar, we used cognitive load theory to develop a self-report instrument to measure three types of cognitive load (intrinsic, extraneous, and germane load), and to provide evidence for instrument validity.

Methods: We developed the instrument (the "CLIC" – Cognitive Load Inventory for Colonoscopy) using a multi-step process. It included 19 items measuring three types of cognitive load, 3 global rating items, and demographics. We then conducted a cross-sectional survey administered electronically to 1,061 gastroenterology trainees in the United States. Participants completed the CLIC following a colonoscopy. The two study phases (exploratory and confirmatory) each enrolled for 10 weeks during the 2014-15 academic year. Exploratory factor analysis determined most parsimonious factor structure; confirmatory factor analysis assessed model fit. Composite measures of intrinsic, extraneous, and germane load were compared across years of training and with global rating items.

Results: 477 (45.0%) invitees participated (116 exploratory and 361 confirmatory study) from 154 (95.1%) training programs. Demographics were similar to US national data. The most parsimonious factor structure included three factors reflecting the three types of cognitive load. Confirmatory factor analysis verified a 3-factor model as best fit. Intrinsic, extraneous, and germane load items had high internal consistency (Cronbach's alpha 0.90, 0.87, 0.96 respectively), and correlated as expected with year in training and global assessment of cognitive load.

Conclusions: The CLIC measures three types of cognitive load during colonoscopy training. Evidence for validity is provided. Although CLIC items relate to colonoscopy, the development process we detail can be used to adapt the instrument for use in other learning settings in medical education.

INTRODUCTION

The distinction between knowledge and skills, or the cognitive and psychomotor domains of learning, has been dominant ever since the work of Benjamin Bloom.[1] This dichotomy may have distracted researchers and practitioners in medical education from examining the cognitive demands learners experience while learning to perform procedures that have prominent psychomotor components. However, procedural tasks are increasingly common in medical education and represent an under-researched area.[2, 3]

Cognitive load theory (CLT), proposed by Sweller and colleagues[4] is a learning theory with promising application for medical education.[4-7] CLT describes working memory as the primary bottleneck for learning.[6] Working memory can process only a small number (approximately 4-7) of units of information at any particular moment (collectively referred to as "cognitive load" [CL]). When the number of informational elements exceeds working memory capacity, new information cannot be processed, limiting learning and performance. CLT describes three types of CL.[5] *Intrinsic load* (IL) arises from the difficulty or complexity of the essential portions of a learning task for a given learner. IL is influenced by prior knowledge and experience. For example, inserting a central venous catheter would present high IL to an intern who has never seen, or learned about, central line insertion, but low IL to a senior resident who has performed the task many times. *Extraneous load* (EL) refers to CL imposed by factors not essential to the learning task, particularly distractions in the learning environment and ineffective instructional techniques. For example, a medical student's EL performing a lumbar puncture in a busy emergency department with frequent interruptions would be higher than the same student's EL performing the lumbar puncture in a quiet clinic room. Finally, *germane load* (GL) is attributable to deliberate cognitive processes activated by the learner for learning. GL is maximized when trainees are given a learning task of appropriate complexity, at the appropriate time, with adequate instruction, and with training in metacognitive skills to maximize learning. For example, a new first-year cardiology fellow with no catheterization experience will experience low GL if she attempts to learn the complexities of coronary artery stent placement, because the IL and EL will consume all her working memory's capacity, leaving no space for GL. The same fellow will encounter higher GL with this same task after learning basic catheterization technique and with careful instruction, because prior knowledge and experience has reduced IL to an appropriate level, and EL is minimized, providing space in working memory for GL. Experts agree that CLT is well-suited to study complex learning settings, such as procedural skills training, where the number of informational elements may exceed the working memory's capacity.[5, 6] CLT has been studied predominantly in classroom settings,[7] with relatively limited examples in medical education.[8-11]

To study the role of CLT in procedural learning, one must measure CL, yet studies measuring CL in medical education settings are limited and not immediately adaptable to diverse procedural settings.[8, 10, 11] Because self-rating instruments have shown promise in accurately measuring the three types of CL in classroom settings,[12, 13] and are efficient and inexpensive to administer, we chose to develop a self-rating instrument to measure CL experienced by learners during procedural training. Considering the great diversity in types and complexity of medical procedures and procedural teaching, it was necessary to focus this study on a single procedure, and we selected colonoscopy. During colonoscopy, a long flexible tube with a camera at the end is used to examine the colon of a patient, who is typically sedated. Biopsies can be taken, and polyps removed, during the procedure.

We selected colonoscopy as an exemplar setting for several reasons. Colonoscopy is a complex learning task and has high cognitive demands (as opposed to more simple medicine-based procedures such as paracentesis). It can be fully accomplished by trainees (as opposed to complex surgical operations), yet entails risks to patients. Colonoscopy is an important procedure for the health of populations. It is one of the most frequently performed medical procedures, and is the preferred method for colorectal cancer screening,[14, 15] yet few studies have critically examined how colonoscopy skills are taught and learned, and how teaching could modify cognitive demands.[16-18]

In this study, we sought to design a self-rating instrument for measuring the CL experienced by gastroenterology trainees (fellows) learning to perform colonoscopy, named the *CLIC* (Cognitive Load Inventory for Colonoscopy), and to provide evidence supporting its validity. We used a rigorous instrument development process that could be used to adapt the CLIC for use in other procedural training settings.

METHODS

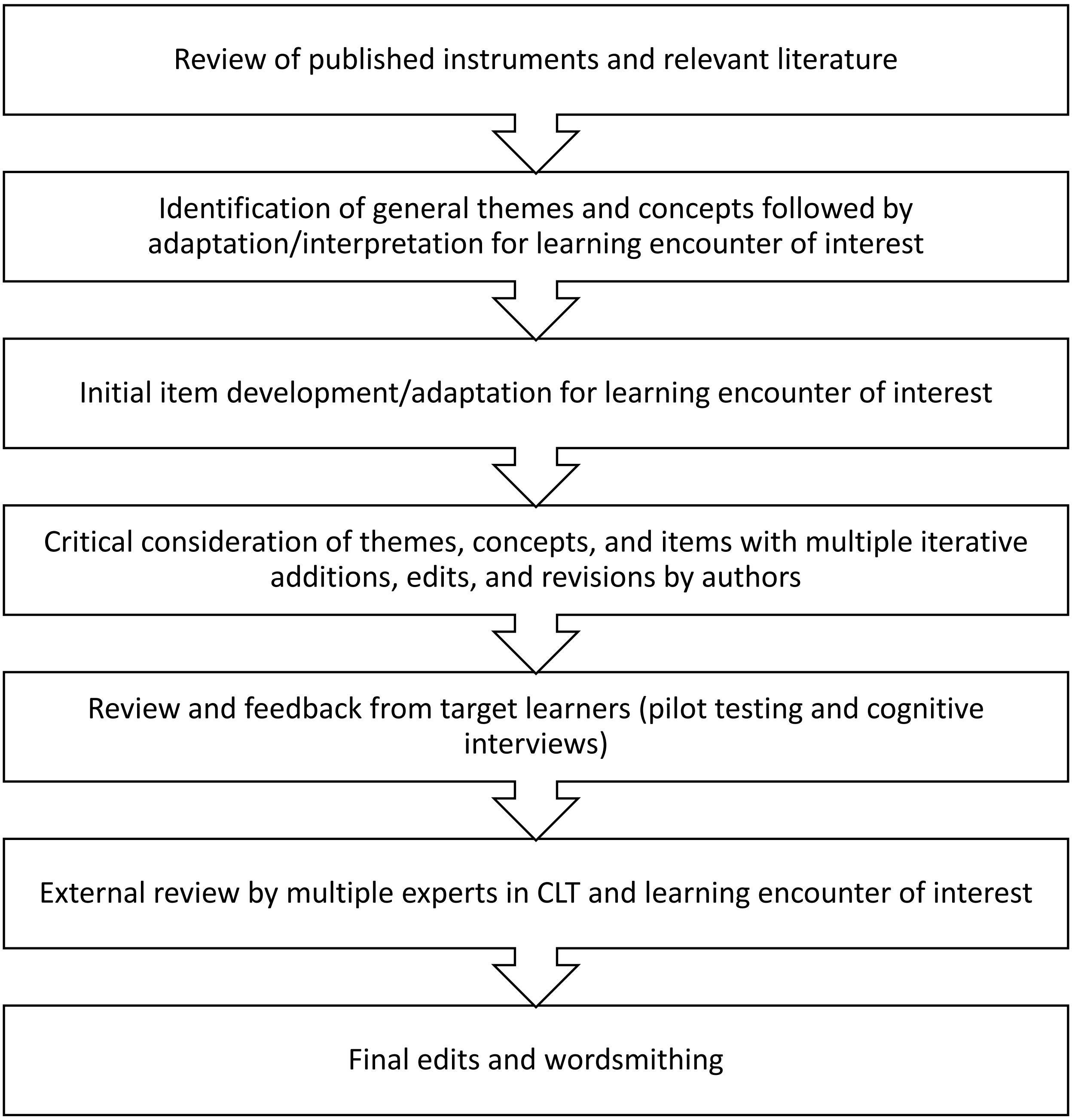
Overall study design

This is a psychometric study based on the unified model of validity[19, 20] designed to provide evidence for validity of a researcher-developed self-rating instrument (CLIC) to measure CL experienced by post-graduate trainees (fellows) in gastroenterology learning to perform colonoscopy. In the exploratory study we identified the most parsimonious factor structure for the CLIC. With the confirmatory study we assessed performance of individual items and model fit. Validity evidence related to content, reliability and association with other variables was collected.

CLIC development

We followed Artino's[21] recommendations to develop the CLIC. Our process is illustrated in Figure 1.

Figure 1. General method for instrument development. In total, there were 31 discrete versions of the instrument (based on the above noted steps and edits by the authors), producing the 19-item instrument tested in our study.



Review of the literature and discussion with gastroenterology fellows and faculty at the University of California San Francisco (UCSF) as well as cognitive load experts within and outside UCSF indicated need for an instrument to measure CL. We searched the literature to identify methods for measuring CL, and identified Leppink's self-rating instrument[13] as a promising starting point given its ability to separately measure IL, EL, and GL. For each of Leppink's items, we asked, "What general learning task-related concept is the item addressing?" For intrinsic and germane items, we interpreted "topic" as overall subject/complexity, "formulas" as motor activities (because formulas are used to "perform" statistics, the setting for Leppink's study), and "concepts and definitions" as cognitive activities. We interpreted extraneous items as relating to content and delivery of instruction by supervisors (teachers). After initial adaptation of these items for colonoscopy, we then asked, "What is missing that could affect IL, EL, and GL during colonoscopy?" This led to further item development to maximize representativeness to CLT and colonoscopy; in particular, we spent significant time crafting wording for GL items. We next pilot-tested the instrument with 16 fellows at UCSF, asking about content, wording, and clarity. Six international experts (three each in colonoscopy and CLT, including Leppink) then provided feedback related to seven specific questions to ensure full coverage of CLT and colonoscopy concepts (notably, the CLT experts agreed with wording of GL items). Finally, we performed cognitive interviews with 8 fellows at UCSF. Revisions were made after each of these steps. The Supplemental Table shows specific source for items and rationale for inclusion.

All authors approved the final version, which included 19 items (Table 1), each employing a 0-10 scale. Anchors were 0 ("strongly disagree") and 10 ("strongly agree"). Items were designed so that a higher response on the scale indicated higher CL. The 11-point scale permitted the ordinal data to approximate continuous data, as in prior studies.[12, 13] Eight items measured IL, 6 items EL, and 5 items GL; responses for each set of items were averaged to calculate total IL, EL, and GL. We developed three global rating items for internal validation (Table 1).

Table 1. Items included in the CLIC with factor loadings, and global rating items.

Item number	Item	Factor loading[1]	Standard error[1]	*t* Test[1]	P-Value[1]
Intrinsic load items: *Please rate your agreement with the following statements regarding the colonoscopy you've just completed:*					
I1	Physically manipulating and controlling the colonoscope was difficult.	0.79	0.024	33.09	<0.001
I2	Using ancillary instruments (i.e., biopsy forceps, polypectomy snare, injector needle, clips, cautery device) was difficult.	0.78	0.024	32.95	<0.001
I3	Identifying normal anatomy and/or landmarks was difficult.	0.72	0.029	24.99	<0.001
I4	The endoscopic findings were complex or difficult to characterize.	0.68	0.032	21.13	<0.001
I5	It was difficult to manage identified pathology (i.e., polyps that were difficult to remove, bleeding that was difficult to control).	0.75	0.027	28.06	<0.001
I6	It was difficult to keep track of, or remember, all the endoscopic findings.	0.69	0.031	22.03	<0.001
I7[2]	Managing the patient's level of comfort was difficult.	---	---	---	---
I8	Overall, this colonoscopy was difficult and/or complex.	0.83	0.020	41.09	<0.001
Extraneous load items: *Please rate your agreement with the following statements regarding your experience during the colonoscopy you've just completed:*					
E1	My supervisor's instructions were unclear.	0.82	0.021	38.83	<0.001
E2	My supervisor used language that was confusing or unfamiliar.	0.90	0.016	57.75	<0.001
E3	The manner in which my supervisor provided instructions or teaching was ineffective for my learning.	0.86	0.018	46.81	<0.001
E4[2]	I felt distracted by other people present in the endoscopy room.	---	---	---	---
E5	I felt distracted by the environment (i.e., my pager going off, environmental noise, the layout of the room).	0.62	0.036	17.41	<0.001
E6[2]	I felt distracted by things on my mind unrelated to this colonoscopy.	---	---	---	---
Germane load items: *Please rate your agreement with the following statements regarding your level of mental effort during the colonoscopy you've just completed:*					
G1	I invested substantial mental effort learning how to control or manipulate the colonoscope and/or other endoscopic equipment.	0.91	0.011	84.84	<0.001
G2	I invested substantial mental effort identifying or understanding colonic anatomy.	0.91	0.011	82.67	<0.001
G3	I invested substantial mental effort understanding, remembering, and/or managing the endoscopic findings.	0.89	0.013	68.79	<0.001
G4[2]	I invested substantial mental effort learning how to manage patient comfort level.	---	---	---	---
G5	Overall, I invested substantial mental effort learning during this colonoscopy.	0.95	0.008	117.90	<0.001
Global rating scales: *During the colonoscopy you've just completed, to what extent was your mind occupied by the following three activities?*					
Global IL[3]	The overall difficulty of this colonoscopy	---	---	---	---
Global EL[3]	Thoughts or distractions not essential to performing or learning colonoscopy	---	---	---	---
Global GL[3]	My efforts to understand and learn colonoscopy technique	---	---	---	---

EL, extraneous load; GL, germane load; IL, intrinsic load. [1] Data relate to confirmatory factor analysis (see results).

[2] These four items were removed during confirmatory factor analysis (see results). [3] Global rating items were not included in factor analysis.

Source of subjects

We recruited gastroenterology fellows in the United States active during the 2014-15 academic year. The American College of Gastroenterology provided a list of email addresses for 913 fellow members. Additionally, we emailed 162 gastroenterology fellowship program directors (identified through websites of the American College of Gastroenterology and the American Gastroenterological Association). Forty-eight responded, providing email addresses for 414 fellows. 190 emails were duplicated between these sources for a final list of 1,137 fellows. In the 2013-14 academic year, there were 1,458 active gastroenterology fellows;[22] assuming similar numbers for 2014-15, we may have identified approximately 80% of active fellows.

Study protocols

Based on review of the literature and expert opinion, it was plausible that our model would be best described using a 3-factor model (including IL, EL, and GL), or a 2-factor model (including only IL and EL, presuming that GL is part of IL, as proposed by some scholars[12]). We performed an exploratory study using exploratory factor analysis (EFA) to determine the most parsimonious factor structure. EFA produced a 3-factor solution, so we proceeded with the confirmatory study, inviting all fellows who had not been invited for the exploratory study; data were analyzed using confirmatory factor analysis (CFA).

Subject recruitment, survey administration and data collection

In keeping with recommendations by Dillman,[23] we contacted invitees multiple times to increase response rate. Participants randomly selected were sent a personal email from a study author (JLS) 1 week before the study began, advising them to expect an invitation by email on a specific date. The CLIC was distributed electronically using REDCap, an academic software program that supports research surveys.[24] Subjects completed the CLIC soon after completing a colonoscopy. Non-responders received weekly reminders over a 10-week period. Invitees could participate only once. Participants could enter a drawing for one of several gift cards.

Data collected

In addition to CLIC items, respondents completed demographic data including year in fellowship, number of previous colonoscopies, gender, geographic region and program size.

Assessment of validity of cognitive load measurement

Consistent with work of Downing and Kane[19, 20] and current standards in educational end psychological measurement,[25] we approached validity assessment unitarily as construct validity, with multiple sources of evidence, and considered potential threats to validity during instrument development.[26] *Content evidence* for validity comes from: adaptation of many items from Leppink's previously published CL measurement instruments;[12, 13] development of additional items to ensure representativeness to the domains of CLT and colonoscopy (our authors have expertise in these domains); developing a larger number of items than we necessarily expected to retain in the final instrument (minimized *construct under-representation*); expert review; and pilot testing and cognitive interviews with fellows (further assessed *item quality*).

Two aspects of our protocol provided evidence supporting validity of *response process*: pilot-testing among fellows at UCSF (identified confusing or poorly-worded questions, reducing *construct-irrelevant variance*); and use of REDCap[24] (ensured quality control of instrument administration using an intuitive, web-based interface, including validation checks).

Evidence for validity of the instrument *internal structure* was provided through factor analysis, which was to confirm that relationships among the items were as hypothesized, specifically that items intended to measure the three types of CL would cluster together and have little overlap with items measuring other CL types. *Internal consistency* was examined as well using Cronbach's alpha. *Generalizability* was enhanced through our national recruitment strategy.

We used two methods to assess *relationships to other variables*. Since prior experience should be the strongest predictor of CL, we measured correlations between year in training and total IL, EL, and GL. Secondly, we correlated total IL, EL, and GL with three global rating items.

Statistical analysis

Because we were primarily concerned with identifying and measuring goodness-of-fit for an anticipated multi-factor solution, we used exploratory followed by confirmatory factor analysis. We treated data as interval data, consistent with expert recommendations,[27, 28] and therefore used parametric methods. We performed EFA using SPSS version 22.0 (IBM Corporation, Armonk, NY), using principle component factor analysis with Varimax rotation. We included all 19 items and selected factors with Eigen values >1.0.

We performed CFA using Mplus 4.0 (Muthen & Muthen, Los Angeles, CA). Following procedures suggested by Hojat,[29] we compared the hypothesized three-factor model with a one- and a two-factor model. We planned *a priori* to then remove items with substantially lower factor loading, taking into account theoretical importance of each item. All models were specified using delta parameterization.[30] Model fit was evaluated using several fit indices: normed χ^2 statistic, comparative fit index (CFI), root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR).[31-33] CFA generates standardized path coefficients for each scale item, which represent the strength of association of each item to the construct (factor) and can be interpreted as correlation coefficients.

Stata version 11.2 (StataCorp, College Station, TX) was used for other analyses. We used one-way ANOVA to assess the three types of CL by year in training, and Pearson's correlation coefficient to measure correlations between the three total CL types and the three global rating items.

Ethical considerations

The Institutional Review Board of the University of California San Francisco reviewed the study protocol and granted it exempt status as an educational research study with minimal risk.

RESULTS

Response rate and characteristics of participants

Of 1,137 potential participants, 27 had undeliverable email addresses, 12 were duplicates with different email addresses, and 37 requested removal resulting in a pool of 1,061 fellows.   We received 477 complete responses (45.0%), representing 154 (95.1%) fellowship programs. Participant demographics were similar in exploratory (N=116) and confirmatory groups (N=361, Table 2). Age and gender were similar to overall US gastroenterology fellows. 1st-year fellows were slightly under-represented.

Table 2. Characteristics of participants.

	Exploratory study (N=116)	Confirmatory study (N=361)	Total (N=477)	ACGME fellows 2013-14 (N=1,458)[1]
Age, mean (SD)	32.2 (3.2)	32.5 (2.8)	32.4 (2.9)	32.3 (NR); P=0.51[2,3]
Female, No. (%)	44 (37.9)	122 (34.1)	166 (35.0)	503 (34%); P=0.90[3]
Year in training, No. (%)				
1st year	25 (21.6)	85 (23.7)	110 (23.2)	486 (33.3); P<0.001[3]
2nd year	53 (45.7)	126 (35.1)	179 (37.7)	482 (33.1); P=0.07[3]
3rd year	34 (29.3)	140 (39.0)	174 (36.6)	490 (33.6); P=0.25[3]
4th year	4 (3.5)	8 (2.2)	12 (2.5)	NR
Prior colonoscopies, No. (%)				
<50	11 (9.5)	14 (3.9)	25 (5.3)	
51-100	14 (12.1)	36 (10.0)	50 (10.5)	
101-150	7 (6.0)	34 (9.5)	41 (8.6)	
151-200	12 (11.2)	36 (10.0)	49 (10.3)	
201-250	16 (13.8)	39 (10.9)	55 (11.6)	
251-300	12 (10.3)	40 (11.1)	52 (11.0)	
>300	43 (37.1)	160 (44.6)	203 (42.7)	
Geographic region, No. (%)				
Northeast	42 (36.2)	134 (37.3)	176 (37.1)	
South	30 (25.9)	83 (23.1)	113 (23.8)	
Midwest	24 (20.7)	76 (21.2)	100 (21.1)	
West	20 (17.2)	66 (18.4)	86 (18.1)	
Number of total fellows in program, No. (%)				
≤6 fellows	30 (25.9)	88 (24.6)	118 (24.9)	
7-12 fellows	61 (52.6)	170 (47.5)	231 (48.7)	
≥13 fellows	25 (21.6)	100 (27.9)	125 (26.4)	

ACGME, Accreditation Council for Graduate Medical Education; NR, not reported

[1]Data are from ACGME Data Resource Book for 2013-14.[27]

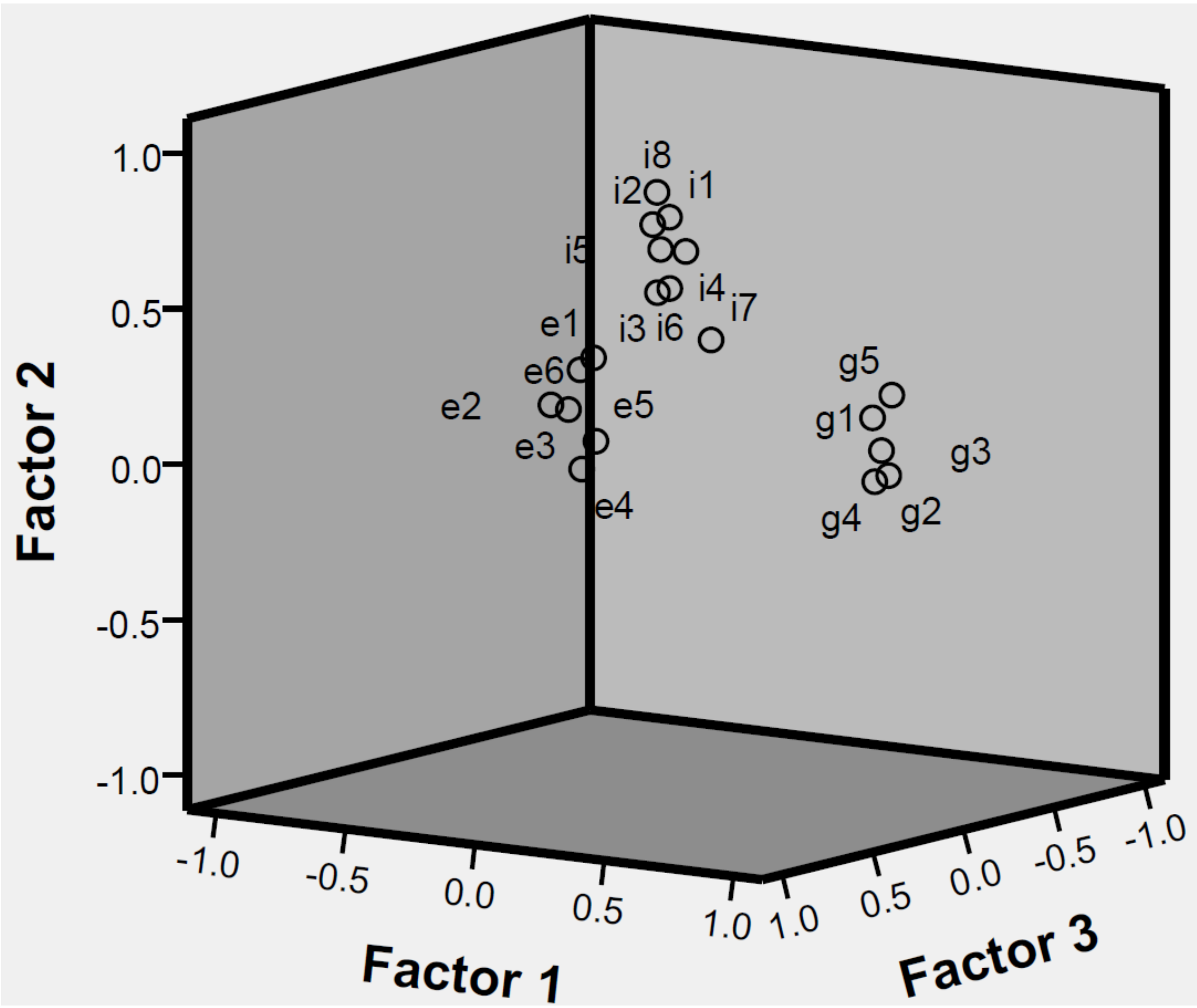
[2] T-test performed assuming that SD for ACGME data is equal to data from our study, as their reference does not cite SD. It is likely that their SD is actually lower given larger sample size.

[3] P-values compare ACGME data with our total population.

Exploratory study

EFA produced a three-factor model explaining 63% of total variance. The factors were consistent with the three types of CL, and all items clustered with like items (Figure 2). Item loadings were high (0.42-0.84 for IL items, 0.56-0.86 for EL items, and 0.80-0.94 for GL items) with very few cross-loadings >0.2.

Figure 2. General method for instrument development. In total, there were 31 discrete versions of the instrument (based on the above noted steps and edits by the authors), producing the 19-item instrument tested in our study.



Confirmatory study

CFA showed superiority of the three-factor model compared with one-factor and two-factor models (Table 3). Three items (I7, E4, E6, Table 1) had notably lower factor loadings (0.67, 0.66, 0.57); these items also had less theoretical support, so we ran a fourth model excluding them (also excluding item G4, because it was linked to I7). This produced the best fit model (normed χ^2 = 4.47, df=87; CFI=0.93, RMSEA=0.098, SRMR=0.04, Table 3). Normed χ^2, CFI, and SRMR were favorable. RMSEA was higher than the more stringent cutoff (0.08) recently recommended, but within the more liberal prior cutoff (0.10)[31]. All factor loadings were well above 0.35 and were statistically significant (Table 1). Internal consistency of the three subscales was high (Cronbach's alpha was 0.90 for IL, 0.87 for EL, 0.96 for GL).

Table 3. Models tested in confirmatory factor analysis with measures of fit.

	χ^2, df, P-value[1] Normed χ^2	CFI[2]	RMSEA (95% CI)[3]	SRMR[4]
One-factor model	χ^2 = 3077.03, df=152, P<0.001 Normed χ^2 = 20.24	0.45	0.231 (0.224,0.238)	0.16
Two-factor model	χ^2 =2090.95, df=151, P<0.001 Normed χ^2 = 13.85	0.63	0.189 (0.182,0.196)	0.20
Three-factor model	χ^2 = 824.02, df=149, P<0.001 Normed χ^2 = 5.53	0.87	0.112 (0.105,0.120)	0.06
Three-factor model excluding items I7, E4, E6, and G4	χ^2 =388.99, df=87, P<0.001 Normed χ^2 = 4.47	0.93	0.098 (0.088,0.108)	0.04

CFI, comparative fit index; CI, confidence interval; df, degrees of freedom; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual

[1] A non-significant (p >0.05) χ^2 suggests the model is an adequate representation of the data. However, with large sample size (>200), χ^2 is almost always significant making the χ^2 fit index inappropriate for larger sample size data such as ours. Given the large sample size, the relative (normed) chi-square is recommended. This value equals the χ^2 index divided by the degrees of freedom. The criterion for acceptance is recommended as less than 5.[38]

[2] CFI is an estimate of the proportion of sample information explained by the model, and can range from 0 to 1; values above 0.90 are generally considered adequate.[37]

[3] RMSEA indicates how well the model fits with the population covariance matrix. Recommended cut-off point for the RMSEA has varied over the years. In the past a value of $\leq$0.1 was considered acceptable, while some scholars have recently proposed a more stringent cutoff of $\leq$0.06.[36]

[4] SRMR is the standardized difference between observed and predicted correlations; a value <0.06 is considered a good fit.[37]

Additional evidence for validity

All three types of CL decreased over time, as expected with advancing expertise (Table 4). Pearson's correlations between measured CL and global ratings (Table 1) were: 0.70 for IL; 0.49 for EL; 0.59 for GL.

Table 4. Three types of cognitive load by year in training.

	First-year fellows (N=110)	Second-year fellows (N=179)	Third- and fourth-year fellows (N=186)	F, P-value
Total intrinsic load, mean (SD)	3.14 (1.78)	1.78 (1.51)	1.21 (1.27)	58.02, <0.0001
Total extraneous load, mean (SD)	1.83 (1.80)	1.14 (1.56)	0.91 (1.32)	12.78, <0.0001
Total germane load, mean (SD)	4.96 (2.12)	3.52 (2.65)	2.85 (2.86)	22.42, <0.0001

DISCUSSION

In this study, we developed and tested an instrument to measure the three types of CL experienced during colonoscopy training among nearly 500 gastroenterology fellows throughout the United States. EFA clearly identified a three-factor structure, with IL, EL and GL items loading into three distinct clusters. CFA confirmed statistical superiority of a three-factor model, consistent with tenets of CLT, with good fit indices. We present evidence supporting a hypothesis of validity for our instrument based on development process and relationships to other variables. Our data support the ability of the CLIC to measure the three types of CL during colonoscopy. The instrument development process we detail could be used to adapt the CLIC for use in other medical education learning settings.

Although CLT is considered highly applicable to medical education,[5] best methods for measuring CL in medical education are not known. We developed a multi-item self-rating instrument because of its potential to measure different types of CL simultaneously, its adaptability for use in diverse learning settings, and its ease and low cost of administration among a large sample of learners. Two recently published studies also addressed methods for measuring CL in medical education settings. The first compared total CL as measured by subjective rating of mental effort and simple reaction time to a vibrotactile stimulus during simple and complex surgical knot-tying simulations among 28 novice medical students.[10] With practice, total CL declined more quickly in the simple versus the complex task group, and there was a trend toward lower subjective rating of mental effort in the simple versus complex task group. The second study compared three measures of CL: Paas' single-item mental-effort rating scale, the NASA Task Load Index, and a 6-item instrument developed by the authors intended to measure the three types of CL among 38 residents undergoing simulation-based procedural training.[11] They found that CL (rescaled to range from 0 to 1) differed across the three measures. There was modest correlation between their 6-item measure and the NASA Task Load Index, but not with the Paas scale. Our adaptable instrument development process, rigorous statistical analysis, large national sample of trainees, ability to separately and distinctly measure the three CL types, and careful collection of evidence for validity builds on these studies and provides support for using multi-item self-rating instruments to measure CL in medical education.

We studied colonoscopy as an exemplar procedure because it represents something of a "middle ground" in procedural training. It is more complex and cognitively demanding than more simple medicine-based procedures (e.g., paracentesis, thoracentesis, central line insertion), yet it is a task that has potential to be completed entirely by a trainee and is relatively structured and protocolized (as opposed to many surgical operations). We believe that the tenets of CLT supported by our study should readily apply to numerous procedural training settings, and that CLT has potential to help address goals and challenges that are consistent across the diverse spectrum of procedural skills training settings in medical education. Procedural skills training universally requires clinician educators to balance learner autonomy and education with patient safety and healthcare quality. We want learners to have as much experience as possible with "hands-on" learning, yet these opportunities are often limited in quantity, and entail risks to patients. CLT's ability to identify and quantify activities of the working memory provides a potential way to address these challenges. If we understand how trainees are using working memory resources, we can identify areas in which CL is imbalanced or overwhelming – areas in which learning is limited and risk of harm to patients increases. In such cases, established methods for optimizing CL[6] have potential to improve both trainee learning and patient safety.

CLT proposes that there are three types of CL that use valuable space in working memory: intrinsic load (task difficulty), extraneous load (distractions, teaching technique), and germane load (deliberate use of learning techniques).[5-7] Experts have suggested based on studies outside of medical education[12, 13] that GL might be part of IL, rather than its own individual type of CL, but our results suggest GL may be separate from IL and EL – an important finding for CLT. It is important to note, however, that we measured trainee *perception* of mental processes, so we cannot prove that GL items truly represent the theoretical version of GL (though we believe this to be the most probable explanation). Since maximal GL is CLT's ideal outcome, better understanding of GL is important. Longitudinal studies linking GL to future performance are needed, as the cognitive schemata and automation occurring through GL should lead to improved performance. Studies in the classroom setting have attempted to address this issue,[12] but the limited length of time between measurement of GL and performance may have been too brief to detect an effect.[34] Should strong correlations between measures of GL and future performance be confirmed, GL could be studied as a surrogate measure for learning, a phenomenon that remains difficult to objectively measure yet is of primary importance in medical education.

This would enable researchers and educators to more immediately measure and understand the effectiveness and impact of individual teachers, learning activities, and curricula, using an understandable and generalizable measure.

As previously noted, we measured fellows' perceptions of their cognitive processes, not their actual cognitive processes in real-time, which was infeasible on such a large scale. Future studies correlating CLIC-measured CL with objective measures of learning (such as task performance) or mental effort (such as functional MRI[35] or psychophysiological measurements[36]) would provide additional evidence for validity and utility of the CLIC. This would also provide further support for teaching procedural skills through the lens of CLT. Adaptation of the CLIC for use in diverse learning settings would also provide additional evidence for validity. With additional future evidence supporting validity of measurement, CL could be a useful and immediately measurable outcome in both research and teaching settings, with potential contributions to curriculum development, program evaluation, and learner assessment.

We enrolled nearly 500 gastroenterology fellows throughout the United States, yet our participation rate was 45%, likely because invitees could participate only after performing a colonoscopy. By contacting 38 program directors (27 responded) we determined that fellows performed colonoscopy on average 8.6 (71.7%) months per year. Reducing our participant pool by this proportion, our response rate could be considered higher (71.7% of 1,061 is 761; 477/761=62.7%). Our participants were representative (Table 2), and we enrolled fellows from 95% of programs in the US. There were fewer 1st-year than 2nd- or 3rd-year fellows in our study, because the American College of Gastroenterology database had fewer 1st-year fellows (20.3%). However, the proportion of 1st-year fellows was similar among invitees (20.4%) and participants (23.2%, P=0.66). Considering the above, we believe our study includes a nationally representative sample of gastroenterology fellows, and illustrates challenges inherent to enrolling a generalizable sample of graduate medical education trainees. We believe that our 10-week enrollment periods and weekly reminders helped improve response rate.

In summary, we developed an instrument that separately measures the three types of CL during an exemplar procedural skills training setting in medical education, among a nationally representative sample of learners. We believe that the CLIC could be readily adapted to measure CL in diverse learning settings in medical education. Our data illustrate the potential impact of systematic instrument development and consideration of factors affecting validity when studying medical education. Teaching trainees to perform medical procedures is a challenging task, considering issues of patient safety, healthcare quality, and time. We believe that CLT provides tools to address these challenges in research and teaching settings.

REFERENCES

1. Bloom B, Engelhart M, Furst E, Hill W, Krathwohl D. *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain.* New York, NY: Longmans & Green; 1956.
2. Mourad M, Ranji S, Sliwka D. A randomized controlled trial of the impact of a teaching procedure service on the training of internal medicine residents. *J Grad Med Educ*. 2012;4(2):170-175.
3. Tukey MH, Wiener RS. The impact of a medical procedure service on patient safety, procedure quality and resident training opportunities. *J Gen Intern Med*. 2014;29(3):485-490.
4. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci*. 1988;12:257-285.
5. Young JQ, van Merriënboer J, Durning S, ten Cate O. Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86. *Med Teach*. 2014;36(5):371-384.
6. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44(1):85-93.
7. van Merriënboer JJ, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev*. 2005;17(2):147-177.
8. Khalil MK, Paas F, Johnson TE, Su YK, Payer AF. Effects of instructional strategies using cross sections on the recognition of anatomical structures in correlated CT and MR images. *Anat Sci Educ*. 2008;1(2):75-83.
9. Qiao YQ, Shen J, Liang X, et al. Using cognitive theory to facilitate medical education. *BMC Med Educ*. 2014;14:79-6920-14-79.
10. Haji FA, Rojas D, Childs R, de Ribaupierre S, Dubrowski A. Measuring cognitive load: performance, mental effort and simulation task complexity. *Med Educ*. 2015;49(8):815-827.
11. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ*. 2015;48(8):805-814.
12. Leppink J, Paas F, van Gog T, ven der Vleuten CPM, van Merriënboer JJ. Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learn Instruct*. 2014;30:32-32-42.
13. Leppink J, Paas F, van der Vleuten CP, van Gog T, van Merriënboer JJ. Development of an instrument for measuring different types of cognitive load. *Behav Res Methods*. 2013;45(4):1058-1072.
14. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L. Association of colonoscopy and death from colorectal cancer. *Ann Intern Med*. 2009;150(1):1-8.

15. Kahi CJ, Imperiale TF, Juliar BE, Rex DK. Effect of screening colonoscopy on colorectal cancer incidence and mortality. *Clin Gastroenterol Hepatol*. 2009;7(7):770-5.
16. Cohen J, Cohen SA, Vora KC, et al. Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. *Gastrointest Endosc*. 2006;64(3):361-368.
17. Spier BJ, Benson M, Pfau PR, Nelligan G, Lucey MR, Gaumnitz EA. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc*. 2010;71(2):319-324.
18. Elta GH. GI training: where are we headed?. *Am J Gastroenterol*. 2011;106(3):395-397.
19. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837.
20. Kane MT. Current concerns in validity theory. *J Educ Meas*. 2001;38(4):319-342.
21. Artino AR,Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach*. 2014;36(6):463-474.
22. Accreditation Council for Graduate Medical Education. *Data Resource Book, Academic Year 2013-2014.* Chicago, IL: ACGME; 2014.
23. Dillman DA, Smyth JD, Christian LM. *Internet, mail, and mixed-mode surveys: the tailored design method.* 3rd ed. New York, NY: John Wiley & Sons, Inc.; 2008.
24. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.
25. Plake B, Wise L, eds. *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association; 2014.
26. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. 2004;38(3):327-333.
27. Sullivan GM, Artino AR,Jr. Analyzing and interpreting data from Likert-type scales. *J Grad Med Educ*. 2013;5(4):541-542.
28. Cook DA, Beckman TJ. Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX. *Adv Health Sci Educ Theory Pract*. 2009;14(5):655-664.
29. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, component, measurement and relationship to gender and specialty. *Am J Psychiatry*. 2002;159(9):1563-1569.

30. Muthen LK, Muthen BO. *Mplus user's guide.* Los Angeles, CA: Muthen & Muthen; 2006.
31. Hooper D, Coughlan J, Mullen MR. Structural Equation Modelling: Guidelines for Determining Model Fit . *The Electronic Journal of Business Research Methods*. 2008;6(1):53-60.
32. Hu L, Bentler PM. Fit indices in covariance structure modelling: sensitivity to under-parameterized model misspecification*. Psychol Methods*. 1998;4(3):424-453.
33. Schumacker RE, Lomax RG. *A beginner's guide to structural equation modeling, 2nd edition.* Mahwah, NJ: Lawrence Erlbaum Associates; 2004.
34. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges*. Perspect Med Educ*. 2015;4(3):107-109.
35. Durning SJ, Costanzo M, Artino AR,Jr, et al. Functional Neuroimaging Correlates of Burnout among Internal Medicine Residents and Faculty Members*. Front Psychiatry*. 2013;4:131.
36. Fredericks TK, Choi SD, Hart J, Butt SE, Mital A. An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads*. Int J Indust Ergonomics*. 2005;35:1097-1107.

SUPPLEMENTAL MATERIAL – CHAPTER 2

Supplemental table. Detailed source and rationale for each item.

Item no.	Item text	Original source	Rationale for inclusion
I1	Physically manipulating and controlling the colonoscope was difficult.	Adapted from Leppink[13] item 2	Motor control of colonoscope contributes to intrinsic task difficulty.
I2	Using ancillary instruments (i.e., biopsy forceps, polypectomy snare, injector needle, clips, cautery device) was difficult.	Adapted from Leppink[13] item 2	Motor control of ancillary items contributes to intrinsic task difficulty.
I3	Identifying normal anatomy and/or landmarks was difficult.	Adapted from Leppink[13] item 3	Cognitive requirement contributes to intrinsic task difficulty.
I4	The endoscopic findings were complex or difficult to characterize.	Adapted from Leppink[13] item 3	Cognitive requirement contributes to intrinsic task difficulty.
I5	It was difficult to manage identified pathology (i.e., polyps that were difficult to remove, bleeding that was difficult to control).	Adapted from Leppink[13] item 2	Motor control to manage pathology contributes to intrinsic task difficulty.
I6	It was difficult to keep track of, or remember, all the endoscopic findings.	Adapted from Leppink[13] item 3	Cognitive requirement contributes to intrinsic task difficulty.
I7	Managing the patient's level of comfort was difficult.	Adapted from Leppink[13] item 3	Managing patient comfort is a cognitive task that contributes to intrinsic task difficulty.
I8	Overall, this colonoscopy was difficult and/or complex.	Adapted from Leppink[13] item 1	General measure of overall task difficulty/complexity.
E1	My supervisor's instructions were unclear.	Adapted from Leppink[13] item 4	Supervisors provide a great deal of instruction during many colonoscopies. If these are unclear, learners use working memory resources to interpret what the supervisor wants them to do.
E2	My supervisor used language that was confusing or unfamiliar.	Adapted from Leppink[13] item 6	Experts often use terms and verbiage with which learners are unfamiliar. Learners use working memory resources as they attempt to determine what this language means.
E3	The manner in which my supervisor provided instructions or teaching was ineffective for my learning.	Adapted from Leppink[13] item 5	Ineffective instructional techniques require learners to use working memory resources to understand and follow instructions.

E4	I felt distracted by other people present in the endoscopy room.	Author generated	Based on authors' and fellows' experience in endoscopy, large numbers of personnel can be distracting. Managing those distractions uses working memory resources.
E5	I felt distracted by the environment (i.e., my pager going off, environmental noise, the layout of the room).	Author generated	Based on authors' and fellows' experience in endoscopy, these environmental occurrences can be distracting. Managing those distractions uses working memory resources.
E6	I felt distracted by things on my mind unrelated to this colonoscopy.	Author generated	Internal distractions can theoretically use working memory resources.
G1	I invested substantial mental effort learning how to control or manipulate the colonoscope and/or other endoscopic equipment.	Related to items I1 and I2	See footnote
G2	I invested substantial mental effort identifying or understanding colonic anatomy.	Related to item I3	See footnote
G3	I invested substantial mental effort understanding, remembering, and/or managing the endoscopic findings.	Related to items I5 and I6	See footnote
G4[2]	I invested substantial mental effort learning how to manage patient comfort level.	Related to item I7	See footnote
G5	Overall, I invested substantial mental effort learning during this colonoscopy.	Related to item I8	See footnote

Footnote: CLT experts agree that germane load is used to manage intrinsic load, through automation and generation of learning schema. We therefore designed our germane load items to assess learners' cognitive processing of the contributors to intrinsic load. This is consistent with Leppink's[13] approach and wording was agreed upon by external CLT expert reviewers.

Chapter 3

Learner, patient, and supervisor features are associated with different types of cognitive load during procedural skills training: implications for teaching and instructional design

Sewell, JL

Boscardin, CK

Young, JQ

ten Cate, O

O'Sullivan, PS

Published as:

Learner, patient, and supervisor features are associated with different types of cognitive load during procedural skills training: implications for teaching and instructional design

Academic Medicine 2017;92(11):1622-1631
DOI: 10.1097/ACM.0000000000001690

ABSTRACT

Purpose: Cognitive load theory, with its focus on limits of the working memory, is relevant to medical education; however, factors associated with cognitive load during procedural skills training are not well-characterized. We sought to determine how features of learners, patients/tasks, settings, and supervisors were associated with three types of cognitive load among learners performing a specific procedure, colonoscopy, to identify implications for procedural teaching.

Method: Data were collected through an electronically administered survey sent to 1,061 US gastroenterology fellows during the 2014-15 academic year; 477 (45.0%) participated. Participants completed the survey immediately following a colonoscopy. Using multivariable linear regression analyses, we identified sets of features associated with intrinsic, extraneous and germane load.

Results: Features associated with intrinsic load included learners (prior experience and year in training negatively associated, fatigue positively associated) and patient/tasks (procedural complexity positively associated, better patient tolerance negatively associated). Features associated with extraneous load included learners (fatigue positively associated), setting (queue order positively associated), and supervisors (supervisor engagement and confidence negatively associated). Only one feature, supervisor engagement, was (positively) associated with germane load.

Conclusions: Our data support practical recommendations for teaching procedural skills through the lens of cognitive load theory. To optimize intrinsic load, level of experience and competence of learners should be balanced with procedural complexity; part-task approaches and scaffolding may be beneficial. To reduce extraneous load, teachers should remain engaged, and factors within the procedural setting that may interfere with learning should be minimized. To optimize germane load, teachers should remain engaged.

INTRODUCTION

Learning to perform medical procedures presents significant challenges to learners and teachers alike.[1,2] This is in part because the learning needs of students, residents, and fellows must be balanced with competing demands of patient safety, procedural efficiency, and healthcare quality. Additionally, learning procedural skills typically includes both cognitive and psychomotor complexity that must be managed simultaneously by learners.

Cognitive load theory (CLT), originally described by Sweller,[3] is a learning theory focused on the working memory, with particular applicability to cognitively complex learning settings, including procedural skills training. CLT holds that the working memory can manage only a few informational elements at any given time under ideal settings; capacity is more limited when learners actively process information. When working memory is overloaded, learning and performance suffer. Such a state has particularly adverse consequences for procedural learning, as the absolute number of opportunities to learn and practice certain procedures may be limited. Procedural learners need to maximize learning from each opportunity. CLT is a useful lens through which to consider procedural training.

CLT envisions learning as occurring when learners use pieces of information to construct a learning "schema," a process also referred to as "chunking".[4] The three types of cognitive load (CL) described by CLT reflect this definition of learning and occur when the working memory is used to manage different activities.[5] *Intrinsic load* (IL) arises from the difficulty of the essential components of a learning task, that is, the activities and thought processes that must be performed to complete a task. Accordingly, IL is highly dependent on learners' knowledge and prior experience. IL will be lower among experts compared with learners, and among advanced learners compared with novice learners. *Extraneous load* (EL) arises from non-essential components of a learning task or learning setting that require learners' attention, such as poor teaching techniques, unclear or confusing instructions, and distractions in the learning environment. Finally, *germane load* (GL) arises from cognitive processes that promote learning, for example, when learners consolidate knowledge, create information chunks, form learning schema, and automate portions of a task.[6] GL is highly dependent on learner effort and metacognitive skills. An ideal learning encounter minimizes EL and optimizes IL to the individual learner's prior experience and level of expertise, which will provide the opportunity for a motivated learner to maximize GL, and, consequently, learning.[7,8] Whether CL is best described by a three-factor (IL, EL, and GL) or a two-factor model (IL and EL only) remains the subject of debate.[7,9,10] In this study, we used a three-factor approach.

CLT is considered relevant to health professions education,[4,6] yet the factors that may affect CL during procedural skills training are not known. Better understanding of such factors could inform development of learning strategies and instructional design for procedural skills training through the lens of CLT. Factors identified as universally affecting CL during procedural skills training could be applied broadly, while factors affecting learning of specific procedural skills could inform teaching strategies for those specific settings.

In this study, we sought to identify features of procedural skills learning encounters associated with intrinsic, extraneous, and germane load. As there is great diversity in the types and complexities of medical procedures and associated learning settings, we felt what was feasible was to focus on the single procedural training setting of colonoscopy. We specifically chose colonoscopy because it represents a middle ground between the simpler, briefer, more focused procedures often performed by non-surgical providers (e.g., intravenous or urinary catheter placement, thoracentesis, paracentesis, arterial or central venous line placement) and the more complex, extended procedures typically performed by surgeons (e.g., full operations). Furthermore, colonoscopy involves substantial cognitive and psychomotor challenges, requiring learners to complete between 300 and 500 cases to attain competence.[11,12] We propose that the study of colonoscopy could provide insights relevant to procedural learning across the spectrum of complexity.

In previous work we reported on development of a self-report instrument (the Cognitive Load Inventory for Colonoscopy, or CLIC) that estimated three types of CL among colonoscopy learners.[13] In this study, we sought to model features of procedural learning encounters that were associated with the three types of CL, as measured by the CLIC, in order to identify implications for instructional design. We conceptualized features as relating to four general aspects of a medical procedure learning encounter: learners, patients/tasks, settings, and supervisors.

METHODS

Study design

We performed a cross-sectional study prospectively enrolling a sample of 1,061 gastroenterology fellows learning to perform colonoscopy throughout the United States during the 2014-15 academic year. Email addresses were obtained through the American College of Gastroenterology and by individually emailing program directors for all gastroenterology fellowship programs in the US. Participants received an email invitation to complete an electronic survey administered via REDCap.[14] They were asked to complete the survey immediately following a colonoscopy. Data collected included cognitive load measures, demographic information, and features of the learning encounter, each of which is described below. 477 (45.0%) invitees participated.

Data for the current study were collected contemporaneously with data from our prior work.[13] That study and this one were conceptualized and planned *a priori* as separate efforts, because they had very different aims and research questions. The aims of the prior study were psychometric – to describe instrument development process and to provide evidence for validity of scores from the CLIC, while the present study aimed to developed models of features associated with the three types of CL.

The Institutional Review Board of the University of California San Francisco reviewed the study protocol and granted it exempt status as an educational research study with minimal risk

Cognitive load measures

IL, EL, and GL were estimated using the CLIC.[13] The final version of the instrument included 15 items; 7 items measured IL, 4 EL, and 4 GL. Items for each CL type were averaged together to calculate composite measures of IL, EL, and GL, which could range from 0 to 10. These IL, EL, and GL measures were the dependent variables of interest for the present study. Supplemental Table 1 depicts items used to measure each type of CL.

Identification and measurement of features that could affect cognitive load

Features that might affect CL during procedural skills training have not been previously studied. Accordingly, through discussion amongst authors, fellows at our training program, and faculty within and outside our training program, we identified features that might contribute to CL during procedural skills training in general, and colonoscopy training in particular. We categorized these features as relating to four general aspects of procedural skills learning encounters: learners, patients/tasks, settings, and supervisors. One feature (supervisor takeover of the procedure) could be considered as relating to multiple groups as the decision to take over a procedure from a learner often takes into account aspects of the learner, the patient/task, the setting, and the supervisor. When available, we utilized categories commonly used in the clinical setting (e.g., bowel preparation quality and patient tolerance of the procedure). As established measurements methods were lacking for many features, we developed our own methods for measurement. Supplemental Table 2 describes how each characteristic was measured and the actual question that participants answered. Two predictor variables (year in training and prior colonoscopy experience) were also reported in our prior work; only year in training was used in analysis for that study.[13]

Demographics

Demographic data included fellow age and gender, geographic region, and training program type and size. Because we did not expect these data to affect CL, they were not included in statistical analyses. These data were also reported in our prior work;[13] we report them here to describe our study sample and provide evidence for adequacy of sampling.

Statistical analysis

Descriptive statistics were calculated using appropriate measures of central tendency and dispersion.

As a research group, we anticipated whether each feature would be expected to have an association (positive or negative) or no association with each type of CL based on our experiences teaching colonoscopy and knowledge about CLT. To reduce risk of type I error, we did not perform univariable statistical analyses.

We developed three multivariable linear regressions models – one each for IL, EL, and GL. We entered as predictor variables the features that we anticipated as associated with each CL type. We considered performing multi-level regression analysis to account for the nested structure of trainees embedded within training programs, but there were inadequate numbers of trainees in most programs to run a robust multi-level model (there were 154 different training programs represented, with number of participants ranging from 1-15 per program, and a median of 2 per program). Our data included six ordinal predictor variables, which were all treated as categorical except for prior colonoscopy experience, as it had 7 response options representing what could be considered continuous data. For variables treated as categorical, when a response category had fewer than 20 responses, it was combined with one or more adjacent response categories to maintain stability and adequate power to estimate the model.

All analyses were performed using Stata version 14.0 (StataCorp, College Station, TX).

RESULTS

Participants

Of the 1,061 fellows invited, 477 (45.0%) participated, representing 154 training programs throughout the United States. Age (mean 32.4 years, SD 2.9) and gender (166 female, 35.0%) of our sample were similar to national data from the Accreditation Council for Graduate Medical Education (mean age 32.3 years [SD not reported], 34% female).[15] Participants came from programs of various sizes (118 [24.9%] with ≤ 6 fellows across all years of training; 231 [48.7%] with 7-12 fellows; 125 [26.4%] with ≥ 13 fellows) and geographic locations in the United States (176 [37.1%] from Northeast; 113 [23.8%] from South; 100 [21.1%] from Midwest; 86 [18.1%] from West). Most participants (457, 96.4%) were training in adult gastroenterology fellowship programs. The above data were also reported in our prior work[13] and are included here only to describe our sample.

Features of learners, patients/tasks, settings, and supervisors

Table 1 shows measures of central tendency and dispersion for features of learners, patients/tasks, settings, and supervisors. Compared with ACGME data,[15] 1st-year fellows were somewhat under-represented. Consistent with year-in-training data, prior colonoscopy experience was skewed toward greater experience, with 203 (42.7%) fellows having previously performed more than 300 colonoscopies.

Screening was the most common indication for colonoscopy (199, 42.3%). Patients were of diverse age ranges; most common was 41-60 years old (N=225, 47.8%). Two hundred thirteen patients (45.3%) were female. Fellows reported most patients to have tolerated the procedure well (370, 78.7%) or fairly well (88 (18.7%), and the quality of most bowel preparations was reported to be excellent (139, 29.5%) or good (253, 53.7%).

Only 15 (3.2%) colonoscopies were performed outside the endoscopy suite. While performing colonoscopy, fellows were generally not on call (53, 11.2%) or paged (105, 22.3%). On average, the colonoscopy on which participants reported was their third procedure performed that day.

Supervisors were usually senior attending physicians (329, 69.4%) or junior attending physicians (138, 29.1%). Fellows typically estimated their supervisors as very engaged (174, 36.5%) or somewhat engaged (194, 40.7%) with teaching, and as very confident (378, 79.2%). The supervisor took over the colonoscopy in a minority of cases (99, 21.0%).

Table 1. Features of learners, patients/tasks, settings, and supervisors among 477 gastroenterology fellows learning to perform colonoscopy during the 2014-15 academic year.

Learner characteristics	N=477
Year in training, No. (%)	
1st year	110 (23.2)[a]
2nd year	179 (37.7)[a]
3rd year	174 (36.6)[a]
4th year	12 (2.5)
Prior colonoscopies, No. (%)	
<50	25 (5.3)
51-100	50 (10.5)
101-150	41 (8.6)
151-200	49 (10.3)
201-250	55 (11.6)
251-300	52 (11.0)
>300	203 (42.7)
Hours of sleep per night during last week, mean (SD)	6.66 (0.85)
Subjective current level of fatigue, 0-100, mean (SD)	38.19 (23.80)
Patient/task characteristics	
Bowel prep quality, No. (%)	
Excellent	139 (29.5)
Good	253 (53.7)
Fair	60 (12.7)
Poor	19 (4.0)
Patient tolerance, No. (%)	
Tolerated well	370 (78.7)
Tolerated fairly well	88 (18.7)
Tolerated somewhat poorly	8 (1.7)
Tolerated poorly	4 (0.9)
Sedation provided by anesthesia, No. (%)	197 (41.7)
Patient age, No. (%)	
<18 years old	6 (1.3)
18-40 years old	47 (10.0)
41-60 years old	225 (47.8)
61-80 years old	174 (36.9)
>80 years old	19 (4.0)
Female patient, No. (%)	213 (45.3)
Number of ancillary maneuvers performed, mean (SD)	1.39 (0.14)
Indication for colonoscopy, No. (%)	
Screening	199 (42.3)
Rectal bleeding	77 (16.4)
Personal history of polyps or cancer	71 (15.1)
Positive fecal occult blood	39 (8.3)
Inflammatory bowel disease	26 (5.5)
Altered bowel habits	19 (4.0)
Abnormal abdominal imaging	13 (2.8)
Anemia	11 (2.3)

Other	15 (3.2)
Urgent or emergent indication for colonoscopy, No. (%)	24 (5.1)
Setting characteristics	
Colonoscopy performed outside the endoscopy suite, No. (%)	15 (3.2)
Queue order, mean (SD)	2.81 (2.33)
Number of people in room, mean (SD)	4.54 (1.07)
On call at the time of colonoscopy, No. (%)	53 (11.2)
Paged during colonoscopy, No. (%)	105 (22.3)
Supervisor characteristics	
Type of supervisor, No. (%)	
Senior attending physician	329 (69.4)
Junior attending physician	138 (29.1)
Another fellow	1 (0.2)
No supervisor	6 (1.3)
Perceived supervisor engagement with teaching, No. (%)	
Very engaged	174 (36.5)
Somewhat engaged	194 (40.7)
Neither engaged nor disengaged	48 (10.1)
Somewhat disengaged	31 (6.5)
Very disengaged	17 (3.6)
Perceived supervisor confidence, No. (%)	4.77 (0.52)
Very confident	378 (79.2)
Somewhat confident	70 (14.7)
Neither confident nor lacking in confidence	15 (3.1)
Somewhat lacking in confidence	2 (0.4)
Very lacking in confidence	0 (0.0)
Supervisor took over colonoscopy, No. (%)	99 (21.0)

[a] In ACGME 2013-14 data,[15] there were 486 (33.3%) 1st-year fellows, 482 (33.1%) 2nd-year fellows, and 490 (33.6%) 3rd-year fellows. Differences compared to our sample were statistically significant for 1st-year fellows (P<0.001), but not for 2nd-year fellows (P=0.07) or 3rd-year fellows (P=0.25).

Predicted associations

Table 2 depicts predicted associations between cognitive load types and features of learners, patients/tasks, settings, and supervisors.

Table 2. Predicted pattern of association between cognitive load types and features of learners, patients/tasks, settings, and supervisors among 477 gastroenterology fellows learning to perform colonoscopy during the 2014-15 academic year.

Features (directionality)	Category	Predicted patterns of association with cognitive load		
		Intrinsic load	Extraneous load	Germane load
Year in training (increasing)	Learner	Negative	Negative	Negative
Prior colonoscopies (increasing)	Learner	Negative	Negative	Negative
Hours of sleep per night (increasing)	Learner	Negative	Negative	None
Level of fatigue (increasing)	Learner	Positive	Higher	Negative
Bowel prep (worse quality)	Patient/Task	Positive	None	None
Patient tolerance (better tolerance)	Patient/Task	Negative	None	None
Sedation by anesthesia	Patient/Task	None	None	None
Female patient	Patient/Task	Positive	None	None
Number of ancillary maneuvers performed	Patient/Task	Positive	None	Positive
Queue order (increasing)	Setting	None	Positive	Negative
Number of people in room (increasing)	Setting	None	Positive	None
On call at the time of colonoscopy	Setting	None	Positive	Negative
Paged during colonoscopy	Setting	None	Positive	Negative
Junior supervisor (versus senior)	Supervisor	None	None	None
Supervisor engagement (less engaged)	Supervisor	None	Positive	Negative
Supervisor confidence (less confident)	Supervisor	None	Positive	Negative
Supervisor took over colonoscopy	Multiple	Positive	Positive	Positive

Multivariable regression analyses

Multivariable linear regression models for each CL type are shown in Table 3. The IL model had an adjusted R^2 value of 0.36 and included six statistically significant features: 3 related to learners, 2 related to patients/tasks, and 1 related to multiple groups. Year in training, prior colonoscopy experience, and better patient tolerance were negatively associated with IL. Fellow fatigue, number of maneuvers performed, and supervisor takeover were positively associated with IL.

The EL model had an adjusted R^2 value of 0.19 and included five statistically significant features: 1 related to learners, 1 related to the setting, 2 related to supervisors, and 1 related to multiple groups. Fellow fatigue, queue order, lower perceived supervisor confidence, lower perceived supervisor engagement, and supervisor taking over the colonoscopy were all positively associated with EL.

The GL model had an adjusted R^2 value of 0.17 and included 1 statistically significant feature. Supervisors perceived as being less engaged were negatively associated with GL. We included estimated IL and EL in the GL model, as these should also be associated with GL. Both were positively associated with GL.

Table 3. Multivariable Linear Regression Models of Features Associated With Intrinsic, Extraneous, and Germane Load Among 477 Gastroenterology Fellows Learning to Perform Colonoscopy During the 2014–15 Academic Year, from a Study of the Implications of Cognitive Load for Teaching and Instructional Design[a]

		Intrinsic load model			Extraneous load model			Germane load model		
Feature	Category	IL coeff[a] (95% CI)	IL β coeff[b]	IL P-value	EL coeff[a] (95% CI)	EL β coeff[b]	EL P-value	GL coeff[a] (95% CI)	GL β coeff[b]	GL P-value
Fellow year in training (year 1 is reference)	Learner									
Year 2	Learner	**-0.60 (-1.086, -0.14)**	**-0.17**	**0.01**	-0.36 (-0.84,0.12)	-0.12	0.14	-0.30 (-1.19,0.60)	-0.05	0.52
Year 3 or 4	Learner	**-0.82 (-1.38, -0.26)**	**-0.23**	**0.004**	-0.32 (-0.89,0.26)	-0.10	0.28	-0.62 (-1.70,0.46)	-0.11	0.26
Prior colonoscopy experience, increasing	Learner	**-0.20 (-0.31, -0.09)**	**-0.24**	**<0.001**	-0.04 (-0.16,0.07)	-0.06	0.46	-0.14 (-0.36,0.07)	-0.10	0.19
Level of fatigue, increasing	Learner	**0.01 (0.008, 0.02)**	**0.21**	**<0.001**	**0.01 (0.005,0.02)**	**0.18**	**0.001**	0.01 (-0.005,0.02)	0.06	0.26
Hours of sleep per night the last week, increasing	Learner	0.11 (-0.07,0.29)	0.05	0.22	0.17 (-0.01,0.35)	0.10	0.07			
Patient tolerated procedure well (compared with fairly well, fairly poorly, or poorly)	Patient/ Task	**-0.67 (-1.01,-0.33)**	**-0.16**	**<0.001**						
Number maneuvers performed, increasing	Patient/ Task	**0.31 (0.18,0.44)**	**0.19**	**<0.001**				-0.06 (-0.32,0.19)	-0.02	0.62
Female patient	Patient/ Task	0.03 (-0.25,0.30)	0.01	0.85						
Bowel prep quality (excellent is reference)	Patient/ Task									
Good	Patient/ Task	0.14 (-0.17,0.44)	0.04	0.38						
Fair or Poor	Patient/ Task	0.17 (-0.26,0.59)	0.04	0.44						
Number of people in room, increasing	Setting				0.04 (-0.09,0.18)	0.03	0.52			
Paged during procedure	Setting				0.16 (-0.19,0.51)	0.04	0.37	0.28 (-0.36,0.93)	0.04	0.39
On call during procedure	Setting				-0.23 (-0.68,0.21)	-0.05	0.30	-0.25 (-1.08,0.58)	-0.03	0.55
Queue order, increasing	Setting				**0.07 (0.01,0.13)**	**0.10**	**0.03**	0.03 (-0.09,0.14)	0.02	0.65
Perceived supervisor engagement (very engaged is reference)	Super-visor									
Somewhat engaged	Super-visor				0.45 (-0.03,0.94)	0.10	0.07	**-0.91 (-1.83,0.001)**	**-0.10**	**0.05**
Neither engaged nor disengaged	Super-visor				0.15 (-0.34,0.64)	0.03	0.55	**-1.33 (-2.24,-0.42)**	**-0.15**	**0.004**
Somewhat disengaged or very disengaged	Super-visor				**0.33 (0.02,0.64)**	**0.11**	**0.04**	**-0.85 (-1.44,-0.26)**	**-0.15**	**0.005**
Perceived supervisor confidence (very confident is reference)	Super-visor									
Somewhat confident, neither confident nor lacking in confidence, or somewhat lacking in confidence	Super-visor				**0.49 (0.13,0.86)**	**0.13**	**0.009**	0.58 (-0.10,1.27)	0.08	0.10
Supervisor took over colonoscopy	Multiple	**0.51 (0.24,0.77)**	**0.16**	**<0.001**	**0.86 (0.49,1.23)**	**0.24**	**<0.001**	-0.43 (-1.16,0.30)	-0.06	0.25
Intrinsic load[c]	N/A							**0.28 (0.08,0.48)**	**0.17**	**0.006**
Extraneous load[c]	N/A							**0.25 (0.06.0.45)**	**0.14**	**0.01**

Abbreviations: CI, confidence interval; coeff, coefficient; EL, extraneous load; GL, germane load; IL, intrinsic load. Bold typeface indicates statistically significant findings. Blank cells indicate that a predictor variable was not included in the model. Thick-lined boxes demarcate categories of predictor variables (horizontal) and the three different regression models (vertical). [a] Unstandardized regression coefficient. [b] Standardized regression coefficient. [c] Intrinsic load and extraneous load were only considered for inclusion in the germane load model.

DISCUSSION

According to CLT, the working memory has limited capacity to manage the three types of cognitive load, intrinsic, extraneous, and germane load, which use working memory resources in an additive fashion.[4] When working memory is overloaded, learning may suffer. It is therefore crucial to understand how learners' working memory resources are being used so cognitive load and learning can be optimized. This is especially pertinent to procedural skills training, where substantial combined cognitive and psychomotor demands may overwhelm working memory and adversely affect learning, and opportunities for deliberate practice may be limited. In this study, we identified features associated with three types of CL among 477 fellows undergoing colonoscopy training in the US. These results suggest practical implications for procedural skills teaching and instructional design through the lens of CLT, which are discussed below, organized by the three types of CL. It is important to note that, to be effective, some of these implications rely on learners and/or instructors being aware of learners' degree of cognitive load, as well as effective communication between learners and instructors.

Intrinsic load

Our findings reflect CLT's stance that IL will be greater when learners are less experienced and/or tasks are more complex (greater number of maneuvers performed, worse patient tolerance of procedure). A less experienced learner and a more complex procedure would both increase the likelihood that the supervisor will need to take over. This was reflected in our model, as supervisor takeover was associated with higher IL.

What do these findings suggest for optimizing IL when teaching procedural skills? While our cross-sectional data prevent direct implication of benefit from any particular teaching approach, there are several that seem relevant based on our data and understanding of CLT. For individual learners, partial task completion and worked examples, both well-studied methods for managing IL,[6] may help balance task complexity with learner experience, particularly among early procedural learners. Considering part-task approaches, an early learner could be assigned only certain components of a procedural task, which could then be expanded (using a scaffolded approach) as the learner advances in experience and level of competence. (For colonoscopy, two competence assessment tools have reported validity evidence and could help inform such decisions.[16,17] Each of these is completed by the supervisor immediately following a colonoscopy.) Considering worked examples, when a very novice learner experiences difficulty, the supervisor could take over immediately (rather than allowing the learner to struggle unproductively), and continuously narrate how s/he is overcoming the difficulty. Ideally, the learner would then be permitted to attempt the technique demonstrated by the supervisor. This might help reduce the IL while promoting GL, but would require the learner and/or the supervisor to recognize when the learner is struggling or is cognitively overloaded.

Our findings might also support benefit of assigning procedural learning tasks that are appropriate for a learner's level of experience. Just as a teacher would not assign an algebra problem to a child early in grade school, nor should we assign an early procedural skills learner a task that is clearly beyond his skills. Such mismatch has been referred to "destructive friction"[18] and puts the learner at substantial risk of cognitive overload. Conversely, a thoughtful and intentional approach to curricular design and individual teaching may help match the procedural task complexity to the skill of the learner.

At the program and curricular level, the classic Halstead approach of "see one, do one, teach one" seems inappropriate for the complexity of modern health professions procedures, as this places learners at risk of overwhelming IL. Rather, a scaffolded approach to procedural learning, such as the "Learn, See, Practice, Prove, Do, Maintain" pedagogical framework proposed by Sawyer and colleagues,[19] may better balance factors contributing to IL.

It is interesting that prior colonoscopy experience and year in training were independently significant in the IL model despite clear correlation with one another. This might suggest that experience outside of actual procedures performed by a learner may contribute to procedural skills learning and competence. Such experiences might include practice with a simulator, observing other learners or faculty performing procedures, discussing aspects of procedural skills during non-procedural patient care, and/or didactic teaching sessions. Additionally, these data may suggest that learners may transfer skills learned from one procedure (e.g., upper gastrointestinal endoscopy) to another procedure (e.g., colonoscopy).

Extraneous load

In the procedural skills setting, EL should arise from distractions in the procedural space, unclear or confusing instructions, or poor teaching techniques, all of which are primarily related to the setting and the teacher. In accord, we found that a setting feature (queue order) and two supervisor features (engagement and confidence) were associated with EL. The lack of association between EL and measures of experience suggests that learners of all stages in training may be susceptible to factors causing EL, and underscores the importance of minimizing EL throughout procedural skills training. Following are suggestions for minimizing EL that could be implicated by our data and CLT.

The number of endoscopies previously performed that day (i.e., queue order) was positively associated with EL, suggesting that, as the day progresses, learners may experience higher EL. Queue order is known to reduce polyp detection rate among practicing gastroenterologists,[20] and it stands to reason this would affect trainees as well. This suggests procedural learners should not perform "too many" procedures per day or session (though the "ideal" number will depend on the setting and the learner), and, that supervisors should carefully select which procedures learners perform near the end of a long day of procedures, or adjust the amount of support they provide to the learner accordingly. Other potentially distracting features of settings, such as pagers, on-call status, and numbers of people in the procedural space were not associated with EL. It may be that these features do not significantly affect EL during colonoscopy training, but could be relevant in other procedural skills settings.

Perceived supervisor disengagement and lack of confidence were associated with higher levels of EL. These intuitive findings suggest that supervisors should remain engaged with the learner and their learning process, and confident in their procedural and teaching skills, to help learners remain on task, and to manage the environment to minimize distractions.

The positive association between EL and supervisor takeover of the colonoscopy might suggest that a learner who is distracted or who receives confusing or unclear instructions (i.e., a learner with high EL) may perform poorly, contributing to supervisor takeover, or, conversely, that supervisor takeover causes learners to perceive more EL. EL should be minimized in procedural teaching to afford learners greater opportunities to complete procedures autonomously.

Germane load

GL involves processes that lead to actual learning as defined by CLT,[4] and features associated with GL could be present in any of the categories that we studied. The CLIC is not definitively proven to measure actual GL, however, as this would require association with measures of learning and/or performance.

In our GL model, only perceived supervisor engagement was associated with GL, and in the anticipated direction (less perceived engagement associated with lower GL). This supports the ability of an engaged supervisor to help learners use cognitive techniques to manage the IL for a procedural task, including schema formation (e.g., how to reduce a colonoscopic loop or remove a polyp), chunking of information (e.g., the anatomical features that together identify the cecum), and task automation (e.g., retroflexion of the colonoscope within the rectum). A supervisor perceived as very engaged might also have been a supervisor who demanded a high level of explanation or performance from the fellow; this would also promote GL. Furthermore, a more engaged supervisor was associated with lower EL, which might have made available more space in working memory for GL. Clearly, supervisors should remain engaged with the procedure and their learner to facilitate GL and learning. However, teachers should also be aware of the expertise reversal effect, in which certain teaching strategies beneficial for novices may actually hamper learning among more experienced learners.[21] As supervisors remain engaged, they must modulate their degree and type of engagement to the learner's degree of expertise.

Why were features of settings, patients/tasks, and learners not associated with GL? It is likely that the setting and patient/task features we assessed have little impact on GL. In other words, procedural skills learners can probably learn from wide variety of patients, tasks, and settings. We suspect that the lack of association with learner features relates to our inclusion of IL in the GL model. Indeed, when we removed IL and EL from the GL model, prior colonoscopy experience and fatigue were both associated with GL, in the same directionality as their associations with IL. Conversely, the IL model did not change with the addition of GL.

The role of fatigue

Fatigue was positively associated with both IL and EL, but not GL. This is an important finding, as trainee work hours and fatigue remain a significant focus of medical education research, particularly in surgical fields, which are intensely procedural.[22] Fatigue's positive association with both IL and EL suggests that fatigue may increase learner perception of intrinsic task difficulty and enhance awareness of, or susceptibility to, environmental distractions and suboptimal teaching technique. However, the lack of association with GL suggests that fatigue may not affect GL (and therefore opportunities for learning). Studies from the surgical literature have found no difference in surgical outcomes for operations performed at night (when fatigue would be higher),[23] or after implementation of work hours restrictions for trainees (which should reduce fatigue).[24] In fact, a recent study of less-restrictive duty hours found non-inferior outcomes in terms of patient safety and trainee satisfaction.[25] While our findings cannot directly support either restricting or liberalizing duty hours for procedural learners, associations between fatigue and IL and EL may provide insights for supervising and teaching procedural learners. It may be beneficial to monitor learners' level of fatigue and to adjust complexity of procedural cases or degree of support for more fatigued learners. This could help avoid high levels of IL and EL, which in some cases could theoretically affect GL and learning.

A summary of the forgoing recommendations to optimize cognitive load during colonoscopy training is provided in Table 4.

Table 4. Teaching and instructional design suggestions to optimize cognitive load during procedural skills training.

Suggestion for teaching or instructional design	Cognitive load goals potentially affected		
	Optimize IL	*Minimize EL*	*Maximize GL*
Have early learners complete only pre-determined part of procedural task	Yes		
Select procedural task based on learner's level and anticipated complexity	Yes		Yes
Take over quickly when early learner is struggling and describe approach	Yes		Yes
Adjust amount of support provided to learner based on learner's prior experience and complexity of procedure	Yes		Yes
Scaffolded procedural skills curriculum	Yes		Yes
Use simulator for early procedural learners		Yes	Yes
Observe others performing procedures	Yes	Yes	Yes
Take advantage of opportunities to learn outside of actual procedures	Yes		Yes
Monitor effects of fatigue on learners	Yes	Yes	
Consider target number of procedures a learner should perform in one session		Yes	
Supervisor manages environment to keep it conducive for learning		Yes	Yes
Supervisor remains engaged with procedure and learner		Yes	Yes
Supervisor keeps learner engaged when taking over procedure			Yes
Ask learners to decide or explain next steps, rather than immediately telling them what to do			Yes

Abbreviations: EL, extraneous load; GL, germane load; IL, intrinsic load.

Limitations of features studied

The features we studied were chosen based on review of the CLT and colonoscopy literature, and after discussions with multiple experts in CLT and colonoscopy training. Although we identified a number of features that were associated with IL, EL, and GL, our models only explained 36%, 19% and 17% of variance, respectively. This begs the question, "What is missing from our models that could affect CL in procedural skills training?" There are undoubtedly unmeasured factors that could affect each type of CL that would in part account for the small to moderate adjusted R^2 values of our models. Because human behavior and psychology (the primary domains of CL) are more complex and less predictable than, for example, physiologic processes, greater variance and lower R^2 values are expected. Furthermore, our models were not intended to predict precise estimates of CL, but, rather, to identify factors associated with each type of CL.

Implications for cognitive load theory

There remains debate regarding GL and whether it represents a unique component of CL,[7] or whether it should be considered as part of IL.[9,10] This debate remains, in part, because few studies have separately measured different CL types.[8,9,26,27] Though the present study cannot fully solve this debate, the differing sets of contributors to IL and GL provide additional evidence supporting a three-factor solution for CL. Characteristics clearly related to intrinsic task difficulty (patient tolerance, number of maneuvers performed, whether the supervisor took over the colonoscopy) were associated with IL but not GL, while perceived supervisor engagement was associated with GL but not IL. If IL and GL represent the same construct, we should have found similar characteristics associated with each. Our findings therefore provide additional support for (but do not prove) a three-factor model for CL.

Limitations

Because studying procedural skills training in all its variety across the spectrum of medical education was infeasible, we selected colonoscopy as an exemplar, for the reasons discussed previously. While some of the predictor variables we studied are specific to colonoscopy, the general categories of variables we studied (learners, patients/tasks, settings, supervisors) are pertinent and adaptable to any procedural skills training setting in the health professions.

The CLIC estimated trainee *perceptions* of CL, and did not measure CL directly. However, only psychometric methods such as ours can currently measure different types of CL in the colonoscopy setting, and the three CL types were central to our research questions. Our prior work provided evidence that the CLIC measures three different constructs, which we believe are likely the three types of CL.[13] The NASA Task Load Index (which measures perceived level of exertion) has been used to study colonoscopy training,[28] yet we felt that studying different types of CL would provide richer implications for instructional design. We acknowledge that the CLIC is a post-hoc assessment of CL. Because self-assessment is considered a relatively unreliable skill among trainees,[29] it is possible that a fellow's perception of how a procedure went could affect their mental reconstruction of the event, their estimation of CL, and their report of some predictor variables (particularly perceived supervisor confidence or engagement). However, assessing three types of CL in the moment during an actual colonoscopy would not likely have been feasible, and certainly not on such a large scale. Comparing in-the-moment and post-hoc assessments of CL would be an interesting direction for future research. Additional evidence for validity of post-hoc psychometric estimation of CL was recently provided by Szulewski and colleagues, whose study suggested that psychometric and physiologic methods could both be used to measure the construct of CL, with strong correlations between these two types of measures.[30]

Finally, it is important to remember this cross-sectional study cannot assign causality to associations between procedural features and CL. Rather, our findings should be viewed as associations and are intended as exploratory analyses to guide future prospective and/or experimental studies, which could be very useful for further understanding practical application of CLT in procedural skills training. Additionally, our study design prevents direct support for the instructional recommendations that we discuss; rather, these represent the opinion of the authors based on data from our study, our understanding of CLT, and the experience of author JS in performing and teaching colonoscopy.

CONCLUSION

In summary, this study highlights how features of learners, patients/tasks, settings, and supervisors are associated with CL in colonoscopy learning. Our findings offer potential teaching implications across the diversity of procedural skills training settings.

REFERENCES

1. Norris TE, Cullison SW, Fihn SD. Teaching procedural skills. *J Gen Intern Med.* 1997;12 Suppl 2:S64-70.
2. Wayne DB, Barsuk JH, McGaghie WC. Procedural training at a crossroads: striking a balance between education, patient safety, and quality. *J Hosp Med.* 2007;2(3):123-125.
3. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci.* 1988;12:257-285.
4. Young JQ, van Merriënboer J, Durning S, ten Cate O. Cognitive Load Theory: Implications for medical education: AMEE Guide No. 86. *Med Teach.* 2014;36(5):371-371-384.
5. Paas F, Tuovinen JE, Tabbers H, van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol.* 2003;38(1):63-71.
6. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010;44(1):85-93.
7. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. *Perspect Med Educ.* 2015;4(3):107-109.
8. Leppink J, Paas F, van der Vleuten CP, van Gog T, van Merriënboer JJ. Development of an instrument for measuring different types of cognitive load. *Behav Res Methods.* 2013;45(4):1058-1072.
9. Leppink J, Paas F, van Gog T, ven der Vleuten CPM, van Merriënboer JJ. Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learn Instruct.* 2014;30:32-32-42.
10. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. *Perspect Med Educ.* 2015;4(3):119-127.
11. Tassios PS, Ladas SD, Grammenos I, Demertzis K, Raptis SA. Acquisition of competence in colonoscopy: the learning curve of trainees. *Endoscopy.* 1999;31(9):702-706.
12. Spier BJ, Benson M, Pfau PR, Nelligan G, Lucey MR, Gaumnitz EA. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc.* 2010;71(2):319-324.
13. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. *Med Educ.* 2016;50(6):682-692.
14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J*

Biomed Inform. 2009;42(2):377-381.
15. Acgme. *Data Resource Book, Academic Year 2013-2014.* ACGME; 2014.
16. Sedlack RE. The Mayo Colonoscopy Skills Assessment Tool: validation of a unique instrument to assess colonoscopy skills in trainees. *Gastrointest Endosc.* 2010;72(6):1125-1133, 1133.e1121-1123.
17. ASGE Training Committee, Sedlack RE, Coyle WJ, et al. ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD. *Gastrointest Endosc.* 2014;79(1):1-7.
18. Vermunt JD, Verloop N. Congruence and Friction between Learning and Teaching. *Learn Instruct.* 1999;9(3):257-280.
19. Sawyer T, White M, Zaveri P, et al. Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med.* 2015;90(8):1025-1033.
20. Lee A, Iskander JM, Gupta N, et al. Queue position in the endoscopic schedule impacts effectiveness of colonoscopy. *Am J Gastroenterol.* 2011;106(8):1457-1465.
21. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. *Educ Psychol.* 2003;38(1):23-31.
22. Ludmerer KM. Redesigning residency education--moving beyond work hours. *The New Engl J Med.* 2010;362(14):1337-1338.
23. Shaw TM, Lonze BE, Feyssa EL, et al. Operative start times and complications after kidney transplantation. *Clin Transplant.* 2012;26(3):E177-183.
24. Scally CP, Ryan AM, Thumma JR, Gauger PG, Dimick JB. Early impact of the 2011 ACGME duty hour regulations on surgical outcomes. *Surgery.* 2015;158(6):1453-1461.
25. Bilimoria KY, Chung JW, Hedges LV, et al. National Cluster-Randomized Trial of Duty-Hour Flexibility in Surgical Training. *N Engl J Med.* 2016;374(8):713-727.
26. DeLeeuw KE, Mayer RE. A comparison of three measures of cognitive load: evidence for separable measures of intrinsic, extraneous, and germane load. *J Educ Psych.* 2008;100(1):223-234.
27. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ.* 2015;48(8):805-814.
28. Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. 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.* 2014;28(3):155-159.
29. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research

agenda. *Acad Med.* 2005;80(10 Suppl):S46-54.

30. Szulewski A, Gegenfurtner A, Howes DW, Sivilotti ML, van Merriënboer JJ. Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. *Adv Health Sci Educ Theory Pract.* 2016.

SUPPLEMENTAL MATERIAL – CHAPTER 3

Supplemental Table 1. Items Used to Measure Cognitive Load Types in the Cognitive Load Inventory for Colonoscopy (CLIC), from a Study of the Implications of Cognitive Load for Teaching and Instructional Design, 2014–2015.

Intrinsic load items *Please rate your agreement with the following statements regarding the colonoscopy you've just completed:*
Physically manipulating and controlling the colonoscope was difficult.
Using ancillary instruments (i.e., biopsy forceps, polypectomy snare, injector needle, clips, cautery device) was difficult.
Identifying normal anatomy and/or landmarks was difficult.
The endoscopic findings were complex or difficult to characterize.
It was difficult to manage identified pathology (i.e., polyps that were difficult to remove, bleeding that was difficult to control).
It was difficult to keep track of, or remember, all the endoscopic findings.
Overall, this colonoscopy was difficult and/or complex.
Extraneous load items *Please rate your agreement with the following statements regarding your experience during the colonoscopy you've just completed:*
My supervisor's instructions were unclear.
My supervisor used language that was confusing or unfamiliar.
The manner in which my supervisor provided instructions or teaching was ineffective for my learning.
I felt distracted by the environment (i.e., my pager going off, environmental noise, the layout of the room).
Germane load items *Please rate your agreement with the following statements regarding your level of mental effort during the colonoscopy you've just completed:*
I invested substantial mental effort learning how to control or manipulate the colonoscope and/or other endoscopic equipment.
I invested substantial mental effort identifying or understanding colonic anatomy.
I invested substantial mental effort understanding, remembering, and/or managing the endoscopic findings.
Overall, I invested substantial mental effort learning during this colonoscopy.

Participants were asked to rate their level of agreement with each item using a 0 ('strongly disagree' to 10 ('strongly agree') scale.

Supplemental Table 2. Measurement Of Learner, Patient/Task, Setting, and Supervisor Features, from a Study of the Implications of Cognitive Load for Teaching and Instructional Design, 2014–2015

Feature	Category	Data type	Survey question and response options
Year in training	Learner	Ordinal	What is your current year in gastroenterology fellowship? *Options: 1, 2, 3, 4*
Prior colonoscopy experience	Learner	Ordinal	Approximately how many colonoscopies have you performed to date? *Options: 50 or fewer, 51-100, 101-150, 151-200, 201-250, 251-300, More than 300*
Sleep	Learner	Continuous	On average, approximately how many hours have you slept each night over the past week? *Options constrained to answers between 0 and 12*
Fatigue	Learner	Ordinal	What is your current level of fatigue? *0-100 slider scale from "as good as it can be" to "as bad as it can be"*
Bowel prep quality	Patient	Ordinal	The patient's bowel prep was: *1-4 scale: Excellent, Good, Fair, Poor*
Poor patient tolerance	Patient	Ordinal	How well did the patient tolerate the colonoscopy? *1-4 scale: Tolerated well, Tolerated fairly well, Tolerated somewhat poorly, Tolerated poorly*
Type of sedation	Patient	Categorical	How was the patient sedated? *Options: Moderate sedation, Monitored anesthesia care, General anesthesia, No sedation*
Patient age	Patient	Ordinal	How old was the patient? *Options: Less than 18 years old, 18-40 years old, 41-60 years old, 61-80 years old, 81 years or older*
Patient gender	Patient	Categorical	What was the patient's gender? *Options: Female, Male*
Number of ancillary maneuvers performed	Patient	Continuous	What ancillary maneuvers did you perform during the colonoscopy (select all that apply): *Options: Biopsy, Forceps polypectomy, Snare polypectomy, Clip placement, Cautery/coagulation, Injection, Foreign body removal, Other*
Indication for colonoscopy	Patient	Categorical	What was the primary indication for the colonoscopy? *For options, see Table 2*
Urgent or emergent indication	Patient	Yes/No	Was the colonoscopy performed for an urgent or emergent indication?
Location colonoscopy performed	Setting	Categorical	In what setting did you perform the colonoscopy?

			Options: Endoscopy suite, Intensive care unit, Operating room, Emergency department, Other
Queue order	Setting	Continuous	Approximately how many endoscopic procedures (both endoscopies and colonoscopies) did you perform today before starting this colonoscopy?
Number of people in room	Setting	Continuous	Approximately how many people were present in the room during the colonoscopy? Please include all people present including yourself, your supervisor, nursing and clinical staff, observers, and the patient.
On call at the time of colonoscopy	Setting	Yes/No	Were you on call at the time of this colonoscopy?
Paged during colonoscopy	Setting	Yes/No	Were you paged during this colonoscopy?
Type of supervisor	Supervisor	Categorical	What was the level of your supervisor for the colonoscopy? *Options: Senior attending physician, Junior attending physician, Another fellow, No supervisor, Other*
Perceived supervisor engagement	Supervisor	Ordinal	How engaged was your supervisor in terms of teaching you during the colonoscopy? *1-5 scale: Very disengaged, Somewhat disengaged, Neither engaged nor disengaged, Somewhat engaged, Very engaged*
Perceived supervisor confidence	Supervisor	Ordinal	How confident did your supervisor seem? *1-5 scale: Very lacking in confidence, Somewhat lacking in confidence, Neither confident nor lacking in confidence, Somewhat confident, Very confident*
Supervisor took over colonoscopy	Supervisor	Yes/No	Did your supervisor physically take over the colonoscopy at any point?

Chapter 4

Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions

Sewell, JL

Maggio, LA

ten Cate, O

van Gog, T

Young, JQ

O'Sullivan, PS

Published as:

Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions
Medical Teacher 2019;41(3):256-270.

ABSTRACT

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. 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. Searching eight databases, we identified 4,571 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 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. 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[1]. Developed by Sweller and colleagues in 1988 [2], CLT focuses on the role of working memory in learning. 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 [3]. 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 [2,3].

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 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 [4,5]. 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) [6]. 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 [1,4,7,8]. 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" [1,7,8]. 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 [4]. Because 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 [4,9]. 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 [10]. 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) [1,3,11]. 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 [1,3,11,12]. 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 [12]. 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). Because study of non-healthcare settings has informed best practices in healthcare (e.g., insights from aviation have shaped the practice of anesthesia [13]), 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 [14]. The Best Evidence in Medical Education (BEME) Collaboration [15] 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 grey 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 [1,3,11,16-23]. 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 [1,3,11,12], 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 March 14, 2016 and rerun July 21, 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.

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

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 [24], 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 [4,9], 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.

Because 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

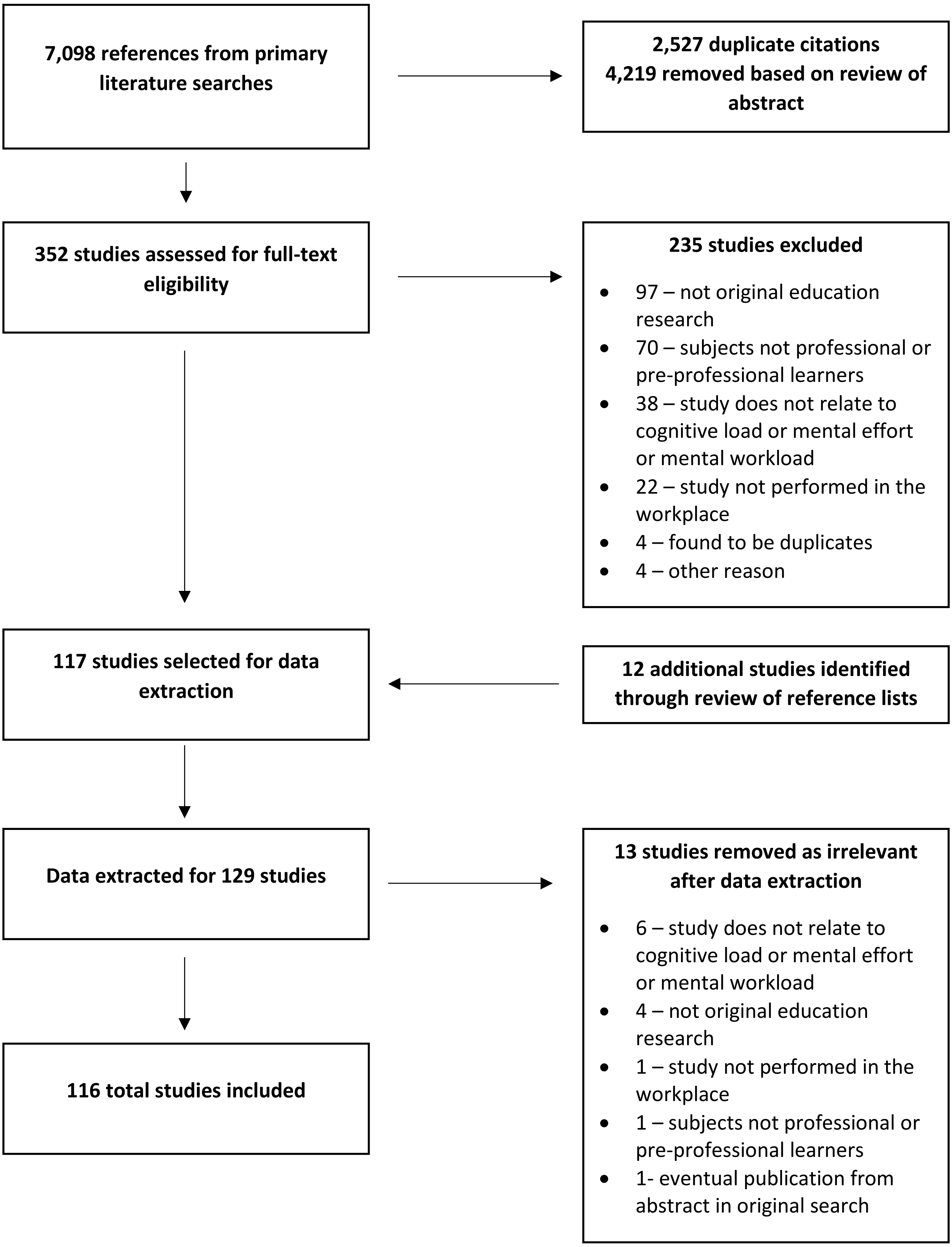
Because of the diverse expertise among our authors, the qualifications of the BEME protocol reviewers, and the anticipated expertise of the editors and scholars who would review our manuscript, we did not separately consult with additional stakeholders. The experiential knowledge of our team was primarily 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 4,571 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 depicting all included studies.

Figure 1. Study selection flow.



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 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.

Table 2. Characteristics of included studies

	Total N = 116
Continent of publication, No. (%)[1]	
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) [1]	
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. (%)[1]	
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. (%)[1]	
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. (%)[1]	
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. (%)[1]	
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)

[1] 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 [25] and element interactivity [26] 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 [10], 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 [27]. A few studies assessed other covariates of CL and performance, including fatigue [28-30], disruptions and distractions [31-35], emotion [17,36-38], the learning environment [39,40], and working with international students [41].

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 [42], laparoscopic gynecological surgical simulations of varying complexity [43], simple versus complex simulated cardiac auscultation [21], and simulated flight [44]. 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 [45]. 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 [46]. In a study involving pilots, a heads-up display designed to reduce pilots' need to derive system states reduced CL and improved performance [47]. 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 [16]. Chen found similar findings related to performance in a study of simulated cardiac auscultation [48], as did Tremblay in a study of simulated medication dispensing and patient counseling among pharmacy students [39]. 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 [49].

Multiple studies assessed how simulated surgical approaches impacted overall CL or intrinsic load: including robotic versus laparoscopic [50-52], simulation versus animal model [53], endoscopic versus laparoscopic versus open [54], and standard versus single incision laparoscopic [55-57]. 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 [42,51-59].

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 [60]. Dahlstrom identified that different flight segments caused varying levels of MWL among trainee and practicing pilots [61]. In a study of ship piloting, Murai found that CL was highest when crucial decisions were required [62]. Similar findings were seen in HPE settings including administration of anesthesia [63,64], clinical reasoning in primary care [65], and creating dental molds [66]. In the latter study, Walker identified which task components were most complex using cognitive task analysis [66].

Although Moos' study suggested that learners may be able to recognize cognitive overload [60], 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 [67] and critical care patient data [68]. 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 [69]. Two studies found that more intuitive formatting of clinical reminders [70] and order sets [71] 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 [65].

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 [72]. In two different studies, Sibbald et al found that use of checklists during simulated cardiac auscultation [73] and simulated electrocardiogram interpretation [74] 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 [39]. 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 [75]. Multi-tasking and time pressure increased CL and impaired nursing students' operation of infusion pumps [76] and surgical residents' laparoscopic suturing [77]. Distractions and disruptions were found to be common sources of extraneous load during medication administration among nurses [35] and among operating room staff [31,32]. Lee and colleagues 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 [78].

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 3 months later was statistically lower than the control group [38]. 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 [36].

Two studies documented teaching behaviors that could impact extraneous load. Tangential conversations were found to increase CL and impair performance during simulated laparoscopic suturing [33]. 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 [40]. 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 [3]. Additionally, 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. Because learning complex skills takes time [16,79-82], 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 [22,83,84]. Two studies of mixed and random practice (for simulated laparoscopic tasks and cardiorespiratory auscultation) found improved performance as compared with blocked practice [21,85], 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 [39], and in simulated emergency medical care among medical students [49]. 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 [86,87]; self-explanation, asking clarifying questions, and/or confirming one's understanding [88]; greater teacher engagement with learners [40]; and careful design of feedback practices [89].

No studies specifically addressed how learners created or honed cognitive schemas. Only one study attempted to link germane load with future performance[49]. 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 [90], patient death [38], vertebroplasty [91], and postoperative care [92]. 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 [92].

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 [93]; testing whether cockpit innovations (such as method or formatting of information display) impacted CL or performance [47,69]; comparing simulated and actual operation of the vehicles [61]; determining the relative levels of CL during particular flight or navigation segments [62,94]; and determining whether increased CL was associated with increased errors [44]. Two studies of military and police personnel examined whether format of training (virtual, live, or mixed) [95], or additional training (situational awareness training) [87] 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 [60,96].

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 Merriënboer and Sweller provided a focused review of CLT as a theory, and specific means to optimize CL in HPE [1]. In 2014 Young (an author on this study) and colleagues summarized CLT tenets as applied broadly to medical education [3]. In 2015, Fraser and colleagues provided a review focused on designing simulation-based learning [11]. In 2015 Naismith and Cavalcanti performed a systematic review assessing CL measurement methods in simulation-based medical education [12]. 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 but 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 [28-30,40]. Fatigue was also associated with extraneous load in one study [40]. 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 [97], 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 [98,99]. Studies have further suggested that, with training, students may themselves be able to select learning tasks appropriate for their zone of proximal development [100]. 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 with in workplaces were found to contribute to extraneous load [65,67-71]. One study also discussed the physical environment as a source of extraneous load [39]. Distractions, disruptions, and interruptions were another common source of extraneous load studied [31-33,35,60]. Additionally, contextual factors, time pressure, and multi-tasking were found as contributing to extraneous load in workplaces [63,76,77,101]. 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 [102]. Indeed, some studies have found that music promotes performance in surgery among novices [103], while others identified negative impact on learners' performance [104]. 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 [70-74,78].

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 [105]? Studies in university students suggest that, with experience, students develop an ability to disregard irrelevant information [106]. 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 [107], 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 [40,49,108,109] than not [88]. 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 [4]. Additional emerging evidence supports separate identities for intrinsic and germane load [110]. 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 dialogue, 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 [111], in which learners must break down part of a learning schema to integrate new information. Because 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 [112], 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 [113-115] 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 [4,7] – 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 Interprofessional 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 [116] and burgeoning interest in variable duration training [117], 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. Additionally, 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 [118]. 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 [98,119], 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 [115] and nuclear power point operation [45], 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 [60,96] 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). Because 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 Table 3. Evidence from included studies is provided when relevant. While some of these recommendations were supported in prior reviews of CLT [1,3] we discuss them here as they impact workplace learning in HPE.

Table 3. Practical suggestions for 'best practices' to optimize cognitive load in health professions workplaces resulting from this review.

Best practices	**Evidence**	**Practical suggestions**
Curricular Design		
Ensure overall CL or intrinsic load of individual learning settings are not *too high*	Performance negatively associated with intrinsic load and overall CL in numerous studies, implying reduced learning when these are too high If intrinsic or overall CL is too high, space for activities contributing to germane load will be limited	• Minimize task complexity for early learners (e.g., partial task, worked examples) [120] • Minimize task complexity when introducing new technologies or methods by providing time, develop a stepwise process • Monitor learners' competence to know when to increase complexity • Scaffold and sequence curriculum to gradually increase complexity over time • Use 4C/ID approach to design individual teaching sessions [121] • Systematically assign patients based on competence [46]
Ensure overall CL or intrinsic load of individual learning settings are not *too low*	Higher complexity tasks associated with higher intrinsic *and* higher germane load among more advanced learners [49] No studies specifically investigated cases in which intrinsic load might be too	• As above

	low, but this could plausibly occur, inducing expertise reversal effect [25]	
Use simulation for early learners, especially when actual tasks involve potential for patient risk and/or have significant element interactivity	Simulation enables control over complexity and deliberate practice [122], which should enhance learning Lower complexity simulation associated with lower CL and better performance (proxy for learning) among novice learners [16,39,48] Studies did not compare simulation with actual patient care	• For procedural skills, build basic skills with simulator, then when competence demonstrated, begin to work with real patients (see: Learn, See, Practice, Prove, Do, Maintain approach [123]) • For cognitive skills, start with simple paper-based cases and then increase complexity of simulation before working with real patients • Monitor lag between training and actual use of skills; may need to plan for practice to bring skills back up to prior level of training
Appraise all aspects of workplaces and associated tasks to identify areas with high potential for cognitive overload	Specific portions of workplace tasks were associated with different levels of CL [60-66]	• Use cognitive task analysis to identify complex areas [27,66] • Orient learners to whole task but allow part task practice until able to accomplish entire task, i.e., 4C/ID [121]
Standardize common tasks, providing supports when needed	Checklists and diagnostic aids reduced CL and improved performance [72-74]	• Use checklists or standard protocols for common, complex, high-risk activities such as patient hand-offs
Design curricula to support workplace learning that	No curricular studies were identified, but indirectly	• Consider how workplace rotation schedule and

scaffold tasks, gradually increasing complexity and reducing support	supported by studies noted earlier in this table, and commentaries suggest benefit [124]	didactic teaching can be used to scaffold learning for critical tasks [120]
Facilitate mixed or random practice over block practice	Lower CL and better performance with mixed or random practice compared with block practice [21,85]	• Integrate skills sessions throughout a curriculum that involve a mixed or random set of cases or diagnoses
Direct Teaching		
Teacher should remain engaged with learning, limiting tangential conversations	Greater teacher engagement associated with higher germane load and lower extraneous load [40] Tangential conversations associated with higher CL, lower performance [33]	• Regularly evaluate teachers' performance and provide constructive feedback • Provide faculty development for teaching engagement techniques, including minimizing tangential conversations which may not over-tax experts but can overload less experienced trainees • Minimize competing duties and distractions for clinical faculty when they are working with learners in the workplace
Teach teachers to monitor for cognitive overload in learners	Although no studies addressed this, teachers are in unique position to observe learners' behavior for cues suggestive of cognitive overload	• Observe learners' body language and utterances for hints of cognitive overload and adjust teaching and level of support accordingly
Attend to learner emotion, especially in crisis situations	Negative emotions negatively affected learning [36,38] Crisis situations were associated with higher levels	• Debrief after crisis situations • Provide faculty development to help

	of CL and lower performance [38,90-92]; debriefing may help [92]	faculty recognize learner emotions and to manage crisis situations and debriefing • When learners display external evidence of negative emotions, discuss with them and facilitate reflection, providing supportive services when needed
Learning Environment		
Leverage graphical displays and technology to reduce extraneous load	Graphical representations of complex data were associated with lower CL and better performance [67-69] Well-designed electronic interfaces (e.g., order sets) were associated with lower CL and better performance [70,71] Consistent with Cognitive Theory of Multimedia Learning [125]	• Critically evaluate tools learners use in clinical settings, optimizing or changing products and interfaces when needed • Thoroughly orient learners to best uses of technology in the workplace – mixed approaches may be most beneficial [95] • Seek learner input for modifiable electronic health record-related tools such as note templates, order sets, and boilerplate phrases
Monitor learning environments for complexity, distractions, and contextual factors that may induce cognitive overload	More complex simulated environments increased extraneous load and decreased performance in novices [39] Contextual factors increased cognitive load and reduced performance, even in experienced physicians [75]	• Critically evaluate workspace and workflow, considering aspects that can be improved to reduce distractions and other sources of extraneous load

Engineer workplace environments to minimize distractions and redundancy	Re-engineered workplace reduced disruptions, lowered CL, and increased satisfaction [78]	• Ask learners what challenges they face in the physical work environment • Involve learners in environment change, possibly as quality improvement effort
Monitor for, and mitigate, learner fatigue	Fatigue was positively associated with extraneous load and overall CL [28,29,40]	• Proactively identify learning settings in which fatigue is common • Adjust tasks, schedules and support to mitigate effects of fatigue on CL
Metacognition		
Help learners know where to direct their limited working memory resources	Specific portions of workplace tasks were associated with different levels of CL [60-66]	• Deconstruct whole tasks into partial tasks, i.e., 4C/ID [121], and prepare learners for those that are likely to be most cognitively demanding
Teach learners to manage distractions	Distractions were present in workplace settings [32,35,82]	• Teach learners that distractions are commonly present in HPE workplaces • Deconstruct with learners typical distractions in the workplace • Simulate distractions and teach skills that help learners manage distractions
Teach learners to monitor their level of CL and	Learners may be able to recognize when they are cognitively overloaded [60]	• Teach learners about CL and provide them language to

communicate feelings of cognitive overload		communicate feelings of overload • Simulate or role-play learners telling their supervisor they feel cognitively overloaded • Provide faculty development so that teachers can respond to learners' cognitive overload
Teach learners to use meta-cognitive techniques to enhance learning	Metacognition was not explicitly studied, but benefits are inferred by tenets of CLT	• Teach learners how to use metacognitive techniques such as self-explanation and monitoring and confirming understanding.

Limitations

We used a scoping methodology [14] 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 detailed above 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 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 [24], 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.

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.

REFERENCES

1. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010;44(1):85-93.
2. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci.* 1988;12:257-285.
3. Young JQ, van Merriënboer J, Durning S, ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Med Teach.* 2014;36(5):371-384.
4. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. *Perspect Med Educ.* 2015;4(3):107-109.
5. Feldon DF. Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educ Psych.* 2007;42(3):123-137.
6. Fiorella L, Mayer RE. Eight Ways to Promote Generative Learning. *Educ Psychol Rev.* 2016;28(4):717-741.
7. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. *Perspect Med Educ.* 2015;4(3):119-127.
8. Kalyuga S. Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educ Psychol Rev.* 2011;23(1):1-19.
9. Leppink J, Paas F, van der Vleuten CP, van Gog T, van Merriënboer JJ. Development of an instrument for measuring different types of cognitive load. *Behav Res Methods.* 2013;45(4):1058-1072.
10. Paas F, Tuovinen JE, Tabbers H, van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol.* 2003;38(1):63-71.
11. Fraser KL, Ayres P, Sweller J. Cognitive Load Theory for the Design of Medical Simulations. *Simulation in healthcare: journal of the Society for Simulation in Healthcare.* 2015;10(5):295-295-307.
12. Naismith LM, Cavalcanti RB. Validity of Cognitive Load Measures in Simulation-Based Training: A Systematic Review. *Acad Med.* 2015;90(11 Suppl):S24-35.
13. Toff NJ. Human factors in anaesthesia: lessons from aviation. *Br J Anaesth.* 2010;105(1):21-25.
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* 2005;8(1):19-32.
15. Thistlethwaite J, Hammick M. The Best Evidence Medical Education (BEME) Collaboration: Into the next decade. *Med Teach.* 2010;32(11):880-882.
16. Haji FA, Rojas D, Childs R, Ribaupierre S, Dubrowski A. Measuring cognitive load: performance, mental effort and simulation task complexity. *Med Educ.* 2015;49(8):815-827.

17. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ.* 2015;48(8):805-814.
18. Szulewski A, Roth N, Howes D. 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.* 2015;90(7):981-987.
19. Khalil MK, Paas F, Johnson TE, Su YK, Payer AF. Effects of instructional strategies using cross sections on the recognition of anatomical structures in correlated CT and MR images. *Anat Sci Educ.* 2008;1(2):75-83.
20. Qiao YQ, Shen J, Liang X, et al. Using cognitive theory to facilitate medical education. *BMC Med Educ.* 2014;14:79.
21. Chen R, Grierson L, Norman G. Manipulation of cognitive load variables and impact on auscultation test performance. *Adv Health Sci Educ Theory Pract.* 2015;20(4):935-952.
22. Andersen SA, Mikkelsen PT, Konge L, Caye-Thomasen P, Sorensen MS. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *Laryngoscope.* 2015;126:E74-E79.
23. Rasmussen SR, Konge L, Mikkelsen PT, Sørensen MS, Andersen SAW. Notes From the Field: Secondary Task Precision for Cognitive Load Estimation During Virtual Reality Surgical Simulation Training. *Eval Health Prof.* 2016;39(1):114-120.
24. Kumasi KD, Charbonneau DH, Walster D. Theory talk in the library science scholarly literature: An exploratory analysis. *Libr Inf Sci Res.* 2013;35:175-180.
25. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. *Educ PSychol.* 2003;38(1):23-31.
26. Hanham J, Leahy W, Sweller J. Cognitive Load Theory, Element Interactivity, and the Testing and Reverse Testing Effects. *Appl Cog Psyc.* 2017;31(3):265-280.
27. Young JQ, ten Cate O, O'Sullivan PS, Irby DM. Unpacking the Complexity of Patient Handoffs Through the Lens of Cognitive Load Theory. *Teach Learn Med.* 2016;28(1):88-96.
28. Bertram DA, Hershey CO, Opila DA, Quirin O. A measure of physician mental work load in internal medicine ambulatory care clinics. *Med Care.* 1990;28(5):458-467.
29. Bertram DA, Opila DA, Brown JL, et al. Measuring physician mental workload: reliability and validity assessment of a brief instrument. *Med Care.* 1992;30(2):95-104.
30. Tomasko JM, Pauli EM, Kunselman AR, Haluck RS. Sleep deprivation increases cognitive workload during simulated surgical tasks. *Am J Surg.* 2012;203(1):37-43.
31. Weigl M, Stefan P, Abhari K, et al. Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-

scale simulated procedure. *Surg Endosc.* 2016;30(2):559-566.
32. Weigl M, Antoniadis S, Chiapponi C, Bruns C, Sevdalis N. The impact of intra-operative interruptions on surgeons' perceived workload: an observational study in elective general and orthopedic surgery. *Surg Endosc.* 2015;29(1):145-153.
33. Gardner AK, Clanton J, Jabbour, II, Scott L, Scott DJ, Russo MA. Impact of seductive details on the acquisition and transfer of laparoscopic suturing skills: Emotionally interesting or cognitively taxing? *Surgery.* 2016;160(3):580-585.
34. Sexton K, Johnson A, Gotsch A, Hussein AA, Cavuoto L, Guru KA. Anticipation, teamwork and cognitive load: chasing efficiency during robot-assisted surgery. *BMJ Qual Saf.* 2018;27(2):148-154.
35. Thomas L, Donohue-Porter P, Stein Fishbein J. Impact of Interruptions, Distractions, and Cognitive Load on Procedure Failures and Medication Administration Errors. *J Nurs Care Qual.* 2017;32(4):309-317.
36. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ.* 2012;46(11):1055-1062.
37. Hautz WE, Schroder T, Dannenberg KA, et al. Shame in Medical Education: A Randomized Study of the Acquisition of Intimate Examination Skills and Its Effect on Subsequent Performance. *Teach Learn Med.* 2017;29(2):196-206.
38. Fraser K, Huffman J, Ma I, et al. The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. *Chest.* 2014;145(5):958-963.
39. Tremblay ML, Lafleur A, Leppink J, Dolmans DH. The simulated clinical environment: Cognitive and emotional impact among undergraduates. *Med Teach.* 2017;39(2):181-187.
40. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. 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.* 2017;92(11):1622-1631.
41. Attrill S, Lincoln M, McAllister S. Supervising international students in clinical placements: perceptions of experiences and factors influencing competency development. *BMC Med Educ.* 2016;16:180.
42. Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen M. Cognitive Load in Mastoidectomy Skills Training: Virtual Reality Simulation and Traditional Dissection Compared. *J Surg Educ.* 2016;73(1):45-50.
43. Bharathan R, Vali S, Setchell T, Miskry T, Darzi A, Aggarwal R. Psychomotor skills and cognitive load training on a virtual reality

laparoscopic simulator for tubal surgery is effective. *Eur J Obstet Gynecol Reprod Biol.* 2013;169(2):347-352.
44. Morris CH, Leung YK. Pilot mental workload: how well do pilots really perform? *Ergonomics.* 2006;49(15):1581-1596.
45. Hsieh MH, Hwang SL, Liu KH, Liang SFM, Chuang CF. A decision support system for identifying abnormal operating procedures in a nuclear power plant. *Nuclear Engineering and Design.* 2012;249:413-418.
46. Young JQ, Niehaus B, Lieu SC, O'Sullivan PS. Improving resident education and patient safety: a method to balance initial caseloads at academic year-end transfer. *Acad Med.* 2010;85(9):1418-1424.
47. Cummings ML, Smith CA. Improving precision vertical flight operations through a direct-perception-action display. *J Aircr.* 2009;46(4):1270-1279.
48. Chen R, Grierson LE, Norman GR. Evaluating the impact of high- and low-fidelity instruction in the development of auscultation skills. *Med Educ.* 2015;49(3):276-285.
49. Dankbaar ME, Alsma J, Jansen EE, van Merriënboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ Theory Pract.* 2016;21(3):505-521.
50. Lee GI, Lee MR, Clanton T, Sutton E, Park AE, Marohn MR. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. *Surg Endosc.* 2014;28(2):456-465.
51. Moore LJ, Wilson MR, McGrath JS, Waine E, Masters RSW, Vine SJ. Surgeons’ display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy. *Surg Endosc.* 2015;29(9):2553-2560.
52. Stefanidis D, Wang F, Korndorffer JR, Jr., Dunne JB, Scott DJ. Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload. *Surg Endosc.* 2010;24(2):377-382.
53. Mouraviev V, Klein M, Schommer E, et al. Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules. *J Robot Surg.* 2016;10(1):49-56.
54. Zheng B, Fu B, Al-Tayeb TA, Hao YF, Qayumi AK. Mastering instruments before operating on a patient: the role of simulation training in tool use skills. *Surg Innov.* 2014;21(6):637-642.
55. Montero PN, Acker CE, Heniford BT, Stefanidis D. Single incision laparoscopic surgery (SILS) is associated with poorer performance and increased

surgeon workload compared with standard laparoscopy. *Am Surg.* 2011;77(1):73-77.
56. Di Stasi LL, Díaz-Piedra C, Ruiz-Rabelo JF, Rieiro H, Sanchez Carrion JM, Catena A. Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: Gaze-based indices. *Appl Ergon.* 2017;65:168-174.
57. Scerbo MW, Britt RC, Stefanidis D. Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. *Am J Surg.* 2017;213(2):244-248.
58. Zheng B, Rieder E, Cassera MA, et al. Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. *Surg Endosc.* 2012;26(5):1352-1358.
59. Britt RC, Scerbo MW, Montano M, Kennedy RA, Prytz E, Stefanidis D. Intracorporeal suturing: Transfer from Fundamentals of Laparoscopic Surgery to cadavers results in substantial increase in mental workload. *Surgery.* 2015;158(5):1428-1433.
60. Moos DC, Pitton D. Student Teacher Challenges: Using the Cognitive Load Theory as an Explanatory Lens. *Teaching Education.* 2014;25(2):127-141.
61. Dahlstrom N, Nahlinder S. Mental workload in aircraft and simulator during basic civil aviation training. *Int J Aviat Psychol.* 2009;19(4):309-325.
62. Murai K, Ishikura A, Hayashi Y, et al. Toward evaluation of ship bridge teamwork for marine simulator training based on heart rate variability. *World Automation Congress.* 2010:1-6.
63. Gaba DM, Lee T. Measuring the workload of the anesthesiologist. *Anesth Analg.* 1990;71(4):354-361.
64. Weinger MB, Vredenburgh AG, Schumann CM, et al. Quantitative description of the workload associated with airway management procedures. *J Clin Anes.* 2000;12(4):273-282.
65. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. *J Gen Intern Med.* 2009;24(3):341-348.
66. Walker J, von Bergmann H. Lessons from a pilot project in cognitive task analysis: the potential role of intermediates in preclinical teaching in dental education. *J Dent Educ.* 2015;79(3):286-294.
67. Doig AK, Albert RW, Syroid ND, Moon S, Agutter JA. Graphical arterial blood gas visualization tool supports rapid and accurate data interpretation. *Comput Inform Nurs.* 2011;29(4 Suppl):Tc53-60.
68. Workman M, Lesser MF, Kim J. An exploratory study of cognitive load in diagnosing patient conditions. *Int J Qual Health Care.* 2007;19(3):127-133.
69. Dominessy ME, Lukas JH, Malkin FJ, Monty RA, Oatman LC. The

effect of information display formats on helicopter pilots' target acquisition and flying performance. *Mil Psychol.* 1991;3(3):163-176.

70. Saleem JJ, Patterson ES, Militello L, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. *J Am Med Inform Assoc.* 2007;14(5):632-640.
71. Avansino J, Leu MG. Effects of CPOE on provider cognitive workload: A randomized crossover trial. *Pediatrics.* 2012;130(3):e547-e552.
72. Liang GF, Lin JT, Hwang SL, Wang EM, Patterson P. Preventing human errors in aviation maintenance using an on-line maintenance assistance platform. *International Journal of Industrial Ergonomics.* 2010;40(3):356-367.
73. Sibbald M, de Bruin AB, Cavalcanti RB, van Merriënboer JJG. Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Qual Saf.* 2013;22(4):333-338.
74. Sibbald M, de Bruin ABH, van Merriënboer JJG. Checklists improve experts' diagnostic decisions. *Med Educ.* 2013;47(3):301-308.
75. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. 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.* 2012;17(1):65-79.
76. Kataoka J, Sasaki M, Kanda K. Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Jpn J Nurs Sci.* 2011;8(1):47-56.
77. Modi H, Leff D, Singh H, Darzi A. The influence of cognitive load on technical ability among surgical trainees. *Int J Surg.* 2016;36:S34-S35.
78. Lee MH, Schemmel AJ, Pooler BD, et al. Radiology Workflow Dynamics: How Workflow Patterns Impact Radiologist Perceptions of Workplace Satisfaction. *Acad Radiol.* 2017;24(4):483-487.
79. Haji FA, Khan R, Regehr G, Drake J, de Ribaupierre S, Dubrowski A. Measuring cognitive load during simulation-based psychomotor skills training: sensitivity of secondary-task performance and subjective ratings. *Adv Health Sci Educ Theory Pract.* 2015;20(5):1237-1253.
80. Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. 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.* 2014;28(3):155-159.
81. Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanström LL. Measuring mental workload during the performance of advanced

laparoscopic tasks. *Surg Endosc.* 2010;24(1):45-50.
82. Haji FA, Cheung JJ, Woods N, Regehr G, de Ribaupierre S, Dubrowski A. Thrive or overload? The effect of task complexity on novices' simulation-based learning. *Med Educ.* 2016;50(9):955-968.
83. Andersen SA, Konge L, Caye-Thomasen P, Sorensen MS. Retention of Mastoidectomy Skills After Virtual Reality Simulation Training. *JAMA Otolaryngol Head Neck Surg.* 2016;142(7):635-640.
84. Donato AA, Kaliyadan AG, Wasser T. Self-directed study using MP3 players to improve auscultation proficiency of physicians: a randomized, controlled trial. *J Contin Educ Health Prof.* 2014;34(2):131-138.
85. Shewokis PA, Shariff FU, Liu Y, Ayaz H, Castellanos A, Lind DS. Acquisition, retention and transfer of simulated laparoscopic tasks using fNIR and a contextual interference paradigm. *Am J Surg.* 2017;213(2):336-345.
86. Saus E, Johnsen BH, Eid J. Perceived learning outcome: the relationship between experience, realism and situation awareness during simulator training. *Int Marit Health.* 2010;62(4):258-264.
87. Saus E, Johnsen BH, Eid J, Riisem PK, Andersen R, Thayer JF. The effect of brief situational awareness training in a police shooting simulator: an experimental study. *Mil Psychol.* 2006;18(Suppl):S3-S21.
88. Young JQ, Boscardin CK, van Dijk SM, et al. Performance of a cognitive load inventory during simulated handoffs: Evidence for validity. *SAGE Open Med.* 2016;4:2050312116682254.
89. Lee GI, Lee MR. 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.* 2018;32(1):62-72.
90. Davis DHJ, Oliver M, Byrne AJ. A novel method of measuring the mental workload of anaesthetists during simulated practice. *Br J Anaesth.* 2009;103(5):665-669.
91. Wucherer P, Stefan P, Abhari K, et al. Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis. *IEEE Trans Med Imaging.* 2015;34(8):1730-1737.
92. Boet S, Sharma B, Pigford AA, Hladkowicz E, Rittenhouse N, Grantcharov T. Debriefing decreases mental workload in surgical crisis: A randomized controlled trial. *Surgery.* 2017;161(5):1215-1220.
93. Crosby JJ, Parkinson SR. A dual task investigation of pilots' skill level. *Ergonomics.* 1979;22(12):1301-1313.
94. Hsu CK, Lin SC, Li WC. Visual movement and mental-workload for pilot performance assessment.

Lec Notes Comput Sci. 2015;9174:356-364.
95. Maxwell DZ, J. Large Scale Testing and Evaluation of Virtual Environments for Infantry Soldier Tasks Comparing Mental Effort for Live Versus Virtual Training Performance Assessments. MODSIM World 2017; April 25-27, 2017; Virginia Beach, VA.
96. Broyles TW, Epler CM, Waknine JW. 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.* 2011;36(1):49-66.
97. Bannert M. Managing Cognitive Load--Recent Trends in Cognitive Load Theory. Commentary. *Learn Instruct.* 2002;12(1):139-146.
98. Camp G, Paas F, Rikers R, van Merriënboer J. Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Comp Hum Behav.* 2001;17(5):575-595.
99. Corbalan G, Kester L, van Merriënboer JJG. Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemp Educ Psychol.* 2008;33(4):733-756.
100. Kostons D, van Gog T, Paas F. Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learn Instruct.* 2012;22(2):121-132.
101. Durning S, Artino AR, Jr., Pangaro L, van der Vleuten CP, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ.* 2011;45(9):927-938.
102. Moris DN, Linos D. Music meets surgery: two sides to the art of "healing". *Surg Endosc.* 2013;27(3):719-723.
103. Siu KC, Suh IH, Mukherjee M, Oleynikov D, Stergiou N. The effect of music on robot-assisted laparoscopic surgical performance. *Surg Innov.* 2010;17(4):306-311.
104. Miskovic D, Rosenthal R, Zingg U, Oertli D, Metzger U, Jancke L. Randomized controlled trial investigating the effect of music on the virtual reality laparoscopic learning performance of novice surgeons. *Surg Endosc.* 2008;22(11):2416-2420.
105. Young JQ, Wachter RM, ten Cate O, O'Sullivan PS, Irby DM. Advancing the next generation of handover research and practice with cognitive load theory. *BMJ Qual Saf.* 2016;25(2):66-70.
106. Rop G, Wermeskerken M, Nooijer JA, Verkoeijen PPJL, Gog T. Task experience as a boundary condition for the negative effects of irrelevant information on learning. *Educ Psychol Rev.* 2018;30(1):229-253.
107. Rotenstein LS, Ramos MA, Torre M, et al. Prevalence of Depression, Depressive Symptoms, and Suicidal

Ideation Among Medical Students: A Systematic Review and Meta-Analysis. *JAMA.* 2016;316(21):2214-2236.
108. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. *Med Educ.* 2016;50(6):682-692.
109. Young JQ, Irby DM, Barilla-LaBarca ML, ten Cate O, O'Sullivan PS. Measuring cognitive load: mixed results from a handover simulation for medical students. *Perspect Med Educ.* 2016;5(1):24-32.
110. Cook DA, Castillo RM, Gas B, Artino AR, Jr. Measuring achievement goal motivation, mindsets and cognitive load: validation of three instruments' scores. *Med Educ.* 2017;51(10):1061-1074.
111. Illeris K. *Contemporary Theories of Learning.* New York, NY: Routledge; 2009.
112. Moreno R. Cognitive load theory: More food for thought. *Instruct Sci.* 2010;38(2):135-141.
113. Guru KA, Esfahani ET, Raza SJ, et al. Cognitive skills assessment during robot-assisted surgery: separating the wheat from the chaff. *BJU Int.* 2015;115(1):166-174.
114. Hussein AA, Shafiei SB, Sharif M, et al. Technical Mentorship during Robot-Assisted Surgery: A Cognitive Analysis? *BJU Int.* 2016;118(3):429-436.
115. Arico P, Borghini G, Di Flumeri G, et al. Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Front Hum Neurosci.* 2016;10:539.
116. Gruppen L, Frank JR, Lockyer J, et al. Toward a research agenda for competency-based medical education. *Med Teach.* 2017;39(6):623-630.
117. Lucey CR, Thibault GE, ten Cate O. Competency-Based, Time-Variable Education in the Health Professions: Crossroads. *Acad Med.* 2018;93(3S Competency-Based, Time-Variable Education in the Health Professions):S1-s5.
118. Lane IF. Professional competencies in health sciences education: from multiple intelligences to the clinic floor. *Adv Health Sci Educ Theory Pract.* 2010;15(1):129-146.
119. Mihalca L, Salden RJCM, Corbalan G, Paas F, Miclea M. Effectiveness of cognitive-load based adaptive instruction in genetics education. *Comp Hum Behav.* 2011;27(1):82-88.
120. Chen HC, O'Sullivan P, Teherani A, Fogh S, Kobashi B, ten Cate O. Sequencing learning experiences to engage different level learners in the workplace: An interview study with excellent clinical teachers. *Med Teach.* 2015;37(12):1090-1097.
121. Vandewaetere M, Manhaeve D, Aertgeerts B, Clarebout G, van Merriënboer JJ, Roex A. 4C/ID in medical education: How to design

an educational program based on whole-task learning: AMEE Guide No. 93. *Med Teach.* 2015;37(1):4-20.

122. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004;79(10 Suppl):S70-81.

123. Sawyer T, White M, Zaveri P, et al. Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med.* 2015;90(8):1025-1033.

124. Leppink J, Duvivier R. Twelve tips for medical curriculum design from a cognitive load theory perspective. *Med Teach.* 2016;38(7):669-674.

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

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

Supplemental table A. Databases and search terms.

Database	Search Terms
PubMed	("cognitive load" OR "cognitive load theory" OR “mental effort” OR “mental load” OR “human channel capacity” OR “mental workload”) AND (“education, professional” OR “professional education” OR education [sh] OR educat* OR training OR trainee* OR students [mesh] OR student* OR learner*)
PsycINFO	("cognitive load" OR "cognitive load theory" OR “mental effort” OR “mental load” OR “human channel capacity” OR “mental workload”) AND (educat* OR education OR training OR trainee* OR student* OR learner*) in ANYWHERE
ERIC	("cognitive load" OR "cognitive load theory" OR “mental effort” OR “mental load” OR “human channel capacity” OR “mental workload”) AND (educat* OR education OR training OR trainee* OR student* OR learner*) in ANYWHERE
CINAHL	("cognitive load" OR "cognitive load theory" OR “mental effort” OR “mental load” OR “human channel capacity” OR “mental workload”) AND (educat* OR education OR training OR trainee* OR student* OR learner*)
Scopus	("cognitive load" OR "cognitive load theory" OR “mental effort” OR “mental load” OR “human channel capacity” OR “mental workload”) AND (educat* OR education OR training OR trainee* OR student* OR learner*) in Article, Title, Abstract, Keywords
Web of Science	("cognitive load" OR "cognitive load theory" OR “mental effort” OR “mental load” OR “human channel capacity” OR “mental workload”) AND (educat* OR education OR training OR trainee* OR student* OR learner*) in TOPIC
IEEE Xplore Digital Library	("cognitive load" OR "cognitive load theory" OR “mental effort” OR “mental load” OR “human channel capacity” OR “mental workload”) AND (educat* OR education OR training OR trainee* OR student* OR learner*)
Google Scholar	allintitle: ("cognitive load" OR "cognitive load theory" OR "mental effort" OR "mental load" OR "human channel capacity" OR "mental workload") AND (education OR training OR trainee)

Supplemental material – Data extraction form.

What is the article ID number? _______________

Who are you?

- o JLS
- o LAM
- o JQY
- o TvG
- o OtC
- o PSO'S

What is the first author's last name? _______________

In what year was the study published? _______________

In which continent(s) was study performed (select all that apply)?

- ☐ Africa
- ☐ Asia
- ☐ Australia
- ☐ Europe
- ☐ North America
- ☐ South America

In which countries was the study performed? _______________

Were study subjects health professions education (HPE) learners, or non-HPE learners?

- o HPE
- o Non-HPE

[If HPE learners]: What profession(s) was/were studied (select all that apply)?

- ☐ Medicine
- ☐ Nursing
- ☐ Pharmacy
- ☐ Dentistry
- ☐ Physical, occupational, or speech therapy
- ☐ Veterinary
- ☐ Other (specify) _______________

[If non-HPE learners] What profession(s) was/were studied? _______________

What level(s) of learners were studied (select all that apply)?

- Pre-professional (i.e., undergraduate)
- Professional (e.g., medical school)
- Post-graduate training
- Practicing professionals (i.e., continuing education)
- Other (specify) _______________
- Can't tell

What was the overall sample size (or note if unclear)? _______________

What was the unit of analysis (if relevant)? _______________

Was/were the primary study aim(s)/purpose(s)?

- Clearly stated
- Implied but not clearly stated
- Could not determine primary study aim

[If could determine study aim] What was/were the primary study aim(s)/purpose(s)? _______________

What is the study setting?

- Workplace
- Simulation - primarily cognitive
- Simulation - primarily psychomotor
- Simulation - combined cognitive and psychomotor
- Other (specify) _______________

What specific activity was studied? _______________

What was the primary study methodology?

- Quantitative
- Qualitative
- Mixed methods
- Experimental
- Other (specify) _______________

[If Qualitative] What was the specific qualitative study design? (Use what author says; add notes in "Other" if questions or disagree with authors).

- o Case study
- o Discourse analysis
- o Ethnography
- o Grounded theory
- o Phenomenology
- o Thematic analysis
- o Other (specify) ________________

[If Quantitative] What was the specific quantitative study design? Select all that apply. (Use what author says; add notes in "Other" if questions or disagree with authors).

- ☐ Descriptive
- ☐ Inferential (including psychometrics)
- ☐ Retrospective
- ☐ Prospective
- ☐ Other (specify) ________________

[If Experimental] What was the specific experimental study design? Select all that apply. (Use what author says; add notes in "Other" if questions or disagree with authors).

- ☐ Randomized controlled trial
- ☐ Pre-test / Post-test
- ☐ Single group, no comparison
- ☐ Single group, repeated measures
- ☐ Other (specify) ________________

What were sources for data (select all that apply)?

- ☐ Surveys (quantitative)
- ☐ Surveys (open ended questions)
- ☐ Objective observations/tests
- ☐ Interviews
- ☐ Focus groups
- ☐ Documents/artifacts
- ☐ Observations with field notes
- ☐ Written texts (narratives, reflections)
- ☐ Other (specify) ________________

Which of the following are mentioned (select all that apply)? *Consider searching in PDF with Ctrl-F.*

- ☐ Cognitive load

- ☐ Intrinsic load
- ☐ Extraneous load
- ☐ Germane load
- ☐ Mental effort
- ☐ Mental workload

Was CLT or mental effort or mental workload measured?

- o Yes
- o No

[If CLT or mental workload was measured] What aspects of cognitive load were measured (select all that apply)? What was the purpose for measurement (i.e., study outcome, covariate, etc.)?

- ☐ Overall cognitive load, mental effort, or mental workload
- ☐ Subtypes of cognitive load (i.e., intrinsic, extraneous, or germane load)
- ☐ Other (specify) ________________

[If CLT or mental workload was measured] What methods were used to measure cognitive load (select all that apply and specify as needed)?

- ☐ Psychometric single item (Paas)
- ☐ Psychometric single item (other)
- ☐ Psychometric multi-item
- ☐ Response time to secondary task
- ☐ Physiologic measures (specify which ones) ________________
- ☐ Other (specify) ________________

How was CLT integrated into the study (based on Kumasi categorization)? (Enter notes afterward if appropriate.)

- o Theory dropping (superficial mention of CLT without citations to primary literature)
- o Theory positioning (limited discussion of CLT with citations to primary work)
- o Theory diversification (CLT and other theories discussed but relevance to study not well-articulated)
- o Theory conversation (CLT discussed in detail and relevance to study is evident, but CLT not integrated into the study design)
- o Theory application (employs CLT throughout to inform research design and data analysis)
- o Theory testing/verification (study designed specifically to test tenets of CLT)
- o Theory generation (seeks to alter or modify CLT)
- o No Other/unsure (specify)

Were primary study outcomes stated? If yes, specify what they were.

- o Yes _______________
- o No
- o Unclear

Was learning and/or performance measured?

- o Yes
- o No
- o Unsure

[If learning and/or performance was measured] How was learning and/or performance measured? _______________

Summarize the primary study findings. _______________

What are primary implications for workplace instruction & learning? _______________

What are primary theoretical implications for CLT? Are findings consistent with, or contradictory to, primary tenets of CLT? _______________

If there were prominent theoretical frames other than CLT, please specify which ones.

____

If there are any reference list that should be pulled for review please list them here.

____

Is this article a potential candidate for discussion as an exemplar?

- o Yes (explain why) _______________
- o No

Should this article be excluded from the review?

- o Yes (explain why) _______________
- o No

If you have any additional comments or notes about this study, please specify here.

____

Supplemental table B. Included studies.

Author and year	Location(s)	Study methodology	Participants[a]	Activity, Simulation or Workplace	Data sources	CL measurement technique(s)[b]	CLT-related study aim(s)[b]	Major finding(s) related to CLT[b]
Andersen 2015[1]	Denmark	Quantitative, experimental	40 "novice" medical students	Mastoidectomy surgery, Simulation	Objective observations or tests	Visual secondary task	To compare CL during virtual reality simulated mastoidectomy training using distributed versus massed practice	Reaction time was lower among those learning mastoidectomy using a distributed versus massed practice approach
Andersen 2016[2]	Denmark	Quantitative, experimental	40 otorhinolaryngology residents	Mastoidectomy surgery, Simulation	Objective observations or tests	Visual secondary task	To compare CL during mastoidectomy training using in traditional dissection versus virtual reality simulation	Reaction time higher in cadaveric than in simulated mastoidectomy
Andersen 2016[3]	Denmark	Quantitative, experimental	36 medical students	Mastoidectomy surgery, Simulation	Objective observations or tests	Visual secondary task	To compare retention of mastoidectomy skills after virtual reality mastoidectomy training in distributed versus massed practice groups and to investigate CL during retention procedures	No difference in reaction time when performing mastoidectomy at transfer after learning from massed versus distributed practice
Andrade 2012[4]	United States	Quantitative, non-experimental	10 "senior" medical students, 10 internal medicine residents, 10 geriatric fellows	Geriatric home safety assessment, Simulation	Quantitative surveys, Interviews	NASA-TLX	To determine whether CL was associated with performance when learning from a 3D geriatric home safety simulation	CL not associated with performance when measured concurrently
Arico 2016[5]	France, Italy	Quantitative, experimental	12 air traffic control students	Air traffic management, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, EEG	To determine whether adjusting task complexity via adaptive automation would impact CL or performance during	Adaptive automation was associated with shorter reaction time and better performance, but subjective CL did not differ

							simulated air traffic management	
Attrill 2016[6]	Australia	Qualitative	20 speech language pathology clinical supervisors	Speech pathology, Workplace	Focus groups	CL not measured	To qualitatively evaluate the experience of supervisors in teaching international students	Teachers felt that international students had overall higher CL than non-international students, particularly related to language issues
Avansino 2012[7]	United States	Quantitative, experimental	3 surgical residents, 3 surgical fellows, 1 surgical attending	Computerized provider order entry, Simulation	Quantitative surveys, Documents/ artifacts	NASA-TLX	To compare CL associated with using a systematically designed order set compared with existing ad hoc order set for simulated entering of post-operative patient orders	Systematically designed order set was easier to use, was associated with lower CL and better performance
Bertram 1990[8]	United States	Quantitative, non-experimental	39 internal medicine residents, 9 internal medicine attendings	Ambulatory patient care, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To develop an instrument to measure CL of physicians working in an internal medicine ambulatory setting, and to identify factors associated with CL	CL positively associated with number of patients seen and fatigue; CL negatively associated with self-rated quality of care and satisfaction with care provided
Bertram 1992[9]	United States	Quantitative, non-experimental	22 internal medicine residents	Ambulatory patient care, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To assess reliability and validity of adapted version of previously developed instrument (Bertram 1990) to measure CL in ambulatory internal medicine setting, and to identify factors associated with CL	CL positively associated with fatigue; CL negatively associated with level of experience, self-rated quality of care provided, and faculty-rated quality of care provided

Bharathan 2013[10]	United Kingdom	Quantitative, non-experimental	25 gynecology residents, 9 gynecology attendings	Laparoscopic gynecologic surgery, Simulation	Objective observations or tests	NASA-TLX, Subjective Mental Effort Questionnaire	To evaluate construct validity for a laparoscopic surgical simulator, to evaluate effectiveness as a training tool, and to examine factors associate with CL	CL was higher for more complex/difficult procedures; Experience positively associated with performance
Blissett 2017[11]	Canada	Quantitative, experimental	86 internal medicine residents	Cardiac auscultation, Simulation	Quantitative surveys, Objective observations or tests	Paas' Subjective Rating of Mental Effort single item scale	To compare how terminally versus branching diagnostic schemas impacted performance and CL during simulated cardiac auscultation	Terminally branching diagnostic schemas (which should require less CL) were associated with lower CL, fewer errors, and greater diagnostic accuracy than hybrid diagnostic schemas
Boet 2017[12]	Canada	Quantitative, experimental	20 surgical residents	Post-operative crisis management, Simulation	Objective observations or tests	Vibrotactile secondary task	To examine whether debriefing after an initial post-operative crisis simulation impacts CL during subsequent simulated post-operative crisis situations	Debriefing after an initial crisis scenario was associated with lower reaction time in a second scenario, compared with a control setting without debriefing
Britt 2015[13]	United States	Quantitative, experimental	14 surgical assistant students	Laparoscopic intracorporeal suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Visual secondary task	To determine whether a new spatial secondary task ("ball-and-tunnel" signal detection task) would identify differences in CL when performing cadaveric versus simulated suturing	Secondary task performance lower and CL higher during cadaveric versus simulated laparoscopic suturing

Broyles 2011[14]	United States	Mixed methods, experimental	27 career and technical education pre-service teachers	Teaching, Workplace	Objective observations or tests, Focus groups	CL not measured	To determine whether degree of CL impacted pre-service teachers' ability to engage in reflection	Higher CL state interfered with ability to engage in reflection
Byrne 1998[15]	United Kingdom	Quantitative, non-experimental	10 anesthesiology residents	Critical anesthesia incident management, Simulation	Objective observations or tests	CL not measured	To examine whether critical incidents during simulated anesthesia negatively impacted accuracy of charting	Accuracy of anesthesia charting was lower during periods of higher CL (simulated patient becomes critically ill), but inaccuracies varied depending on nature of critical incident (e.g., during pulmonary crisis, carbon dioxide level charted more accurately than heart rate)
Byrne 2013[16]	United Kingdom	Quantitative, non-experimental	20 anesthesiology residents and attendings	Administering anesthesia, Workplace	Objective observations or tests	Vibrotactile secondary task	To characterize inter- and intra-person variation in CL during actual administration of anesthesia	Greater variation in reaction time within subjects than between subjects
Cao 2007[17]	United States	Quantitative, experimental	24 surgical residents, 6 surgical fellows	Surgical ball transfer task, Simulation	Objective observations or tests	CL not measured	To determine whether adding haptic feedback during simulated surgical tasks mitigated performance decline associated with cognitively loading subjects	Cognitive loading reduced speed, but not accuracy, on surgical ball transfer task; Adding haptic feedback improved speed and accuracy when cognitively loaded; More experienced surgeons were faster and more accurate
Chen 2015[18]	Canada	Quantitative, experimental	60 "senior-level" nursing students	Cardiac and pulmonary	Objective observations or tests	CL not measured	To compare impact of low- versus high-fidelity instructional techniques	Lower fidelity simulation associated with better performance and recall

				auscultation, Simulation			on performance during simulated cardiac and pulmonary auscultation	than high-fidelity simulation; Both simulations associated with better performance than control group
Chen 2015[19]	Canada	Quantitative, experimental	77 "senior-level" nursing students	Cardiac and pulmonary auscultation, Simulation	Objective observations or tests	CL not measured	To determine whether adjustments intended to impact intrinsic, extraneous, and germane cognitive load had any impact during simulated cardiac and pulmonary auscultation	Less complex task (lower intrinsic load) associated with better performance; Mixed (versus blocked) teaching associated with better performance (implying greater germane load); Authors also concluded that multiple versus single representations of heart sounds did not impact extraneous load
Chowriappa 2015[20]	United Kingdom, United States	Quantitative, experimental	22 surgical residents, 30 surgical fellows	Robotic surgery tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether an augmented-reality-based training ("Hands-on Surgical Training") for simulated robotic assisted surgery impacted CL and performance compared with control condition	Intervention group performed better and had lower CL compared with control group
Chui 2014[21]	United States	Qualitative	8 practicing pharmacists	Pharmacy handoffs, Simulation	Interviews	CL not measured	To describe and characterize "information hazards" present during pharmacy handovers, including whether these	"Information hazards" present during pharmacy handovers included those related to CL: information overload (i.e., intrinsic load) and information

							hazards were related to types of CL	scatter (i.e., extraneous load)
Crosby 1979[22]	United States	Quantitative, experimental	Experiment 1: 12 student pilots, 12 pilot instructors Experiment 2: 11 student pilots	Flight, Simulation	Objective observations or tests	Visual secondary task	To determine whether differences in CL in single- versus dual-task conditions were impacted by student versus instructor status during simulated flight	Differences in response time in single- versus dual-task condition were greater for students than for instructors; Instructors performed better than students; Later-phase students performed similar to instructors
Cummings 2009[23]	United States	Quantitative, experimental	31 student, recreational, and professional pilots	Flight, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether a heads-up informational display impacted performance and CL during simulated flight	No difference in CL or performance when using heads-up informational display during simulated flight
Dahlstrom 2009[24]	Sweden	Quantitative, experimental	11 student pilots, 8 pilot instructors	Flight, Simulation and Workplace	Objective observations or tests	Single item subjective rating of CL, Heart rate variability, Eye movement	To determine whether CL differed in simulated versus actual flight, and in particular flight segments	Neither subjective CL nor overall heart rate variability differed in in simulated compared with actual flight; Heart rate variability lower for some flight segments, suggesting different degree of CL was required
Dankbaar 2016[25]	Netherlands	Quantitative, experimental	79 4th-year medical students	Emergency medical care, Simulation	Quantitative surveys, Objective observations or tests	Investigator-developed multi-item instrument	To compare intrinsic and germane load experienced by medical students during simulated emergency medical care when instructional groups offered differing levels	Comparing control group (e-module only) with two intervention groups (e-module with low fidelity text-based simulations, e-module with high-fidelity simulation game): intrinsic load of e-

							of fidelity and complexity	module was lowest of any element; game group had higher intrinsic load, germane load, and engagement; Performance on skills assessment was same among three groups
Davis 2009[26]	UK	Quantitative, non-experimental	10 anesthesiology residents	Anesthesia crisis management, Simulation	Objective observations or tests	Vibrotactile secondary task	To pilot the use of a "new secondary task measure of CL" (vibrotactile device) in simulated anesthesia crisis management	During crisis portion of anesthesia simulation, response time increased and then went back to baseline
de Man 2014[27]	Netherlands	Quantitative, experimental	17 anesthesiology residents, 10 anesthesiology attendings	Administering anesthesia, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether an "alarm-rich" simulated anesthesia environment would be associated with higher CL and reduced performance (time to detection of adverse events)	No difference in CL performance in alarm-free and alarm-rich simulations
DiStasi 2017[28]	Spain	Quantitative, experimental	8 surgical residents, 8 surgical attendings	Laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Gaze entropy and velocity	To compare cognitive demand of single- versus multi-site simulated laparoscopic surgery	CL (as measured by both gaze entropy and NASA-TLX) was greater with single- compared with multi-port approach; CL as measured by NASA-TLX was higher among residents compared with attendings
Doig 2011[29]	United States	Quantitative, experimental	15 nursing students, 15 critical care nurses	Arterial blood gas interpretation, Simulation	Quantitative surveys, Objective	NASA-TLX	To create and test a computer-aided tool to improve performance	Accuracy and performance were better using tool

					observations or tests		and reduce CL associated during simulated arterial blood gas interpretation	compared with control group; CL was lower using the tool among nursing students, but not among practicing nurses
Dominessy 1991[30]	United States	Quantitative, experimental	16 helicopter pilots	Helicopter flight, Simulation	Objective observations or tests	Reaction time to visual instructions, NASA-TLX	To compare CL during simulated helicopter flight when tactical information was presented in three formats: text, graphic, and numeric, and to determine whether performing an additional task (target acquisition search) impacted performance and CL	CL was highest for text and lowest for graphic; Secondary task reduced performance and increased CL
Donato 2014[31]	United States	Quantitative, experimental	26 residents and 57 practicing physicians (all non-cardiologists)	Cardiac auscultation, Simulation	Quantitative surveys, Objective observations or tests	CL not measured	To compare CL and performance in simulated cardiac auscultation for groups trained with a “lower CL” self-study group or a “higher CL” multimedia lecture group	Participants in the “lower CL” self-study group improved more than the “higher CL” multimedia lecture group
Durning 2012[32]	Netherlands, United States	Mixed methods, non-experimental	25 attending internists	Clinical reasoning, Simulation	Quantitative surveys, Surveys (open-ended questions), Interviews	Utterances and pauses were used to infer CL	To assess how contextual factors anticipated to increase CL (non-English-speaking patient, emotionally volatile patient, diagnostic suggestion, atypical presentation) impact clinical reasoning performance	Contextual factors anticipated to increased CL were associated with poorer diagnostic reasoning performance

Durning 2011[33]	United States	Qualitative	25 attending internists	Patient interviewing, Simulation	Surveys (open-ended questions), Interviews	CL not measured	To qualitatively examine how contextual factors were felt to potentially impact clinical reasoning and CL	Three primary themes arose: (1) Components influencing impact of contextual factors, (2) Mechanisms for addressing contextual factors, and (3) Consequences of contextual factors for patient care; These factors were also felt likely to impact CL
Fenik 2013[34]	Germany	Quantitative, experimental	14 final-year medical students, 16 medical residents	Central line insertion, Simulation	Objective observations or tests	CL not measured	To examine whether use of an "all-inclusive" central line kit during simulated central line insertion would reduce mistakes and improve performance compared with control condition in which subject has to retrieve all supplies separately	The intervention group made fewer mistakes and met more checklist items (better performance), and required less time to complete the insertion, than the control group
Flindall 2015[35]	UK	Quantitative, experimental	20 medical students	Patient handoff, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare deterioration in recall of clinical information 2 hours after simulated patient handoffs for medical students who had completed a "mentally fatiguing high cognitive load task" compared to a control condition	High cognitive loading associated with reduced information freely recalled about cases

Fraser 2012[36]	Canada	Quantitative, non-experimental	84 1st-year medical students	Heart sound auscultation, Simulation	Quantitative surveys, Objective observations or tests	Paas' Subjective Rating of Mental Effort single item scale	To determine whether emotion impacts CL during simulated cardiac auscultation, or post-training performance	Invigoration was positively correlated with CL, and tranquility negatively correlated; Higher CL associated with lower odds of correctly diagnosing a cardiac murmur previously learned, but no impact on diagnosing novel cardiac murmur
Fraser 2014[37]	Canada	Quantitative, experimental	116 final-year medical students	Aspirin overdose management, Simulation	Quantitative surveys, Objective observations or tests	Paas' Subjective Rating of Mental Effort single item scale	To determine the impact of unexpected patient death during simulated aspirin overdose management on CL during simulation, and performance on an observed structured clinical encounter 3 months later	Emotions were more negative and CL higher when the simulated patient unexpectedly died; performance 3 months later was lower in the group in which simulated patient died
Gaba 1990[38]	United States	Quantitative, non-experimental	9 anesthesiology residents	Administering anesthesia, Workplace	Objective observations or tests	Investigator-developed multi-item instrument, Mathematics problem secondary task	To examine the use of a secondary task methodology (solving simple mathematics problems) to assess workload of anesthesiology residents during actual administration of anesthesia	Response times were higher during period requiring subjects to perform manual tasks; Response times did not correlate with level of training or case complexity
Gardner 2016[39]	United States	Quantitative, experimental	41 1st-year, 37 2nd-year, 9 3rd-year, and 2 4th-year medical students	Laparoscopic suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine how "seductive details" (interesting but tangential bits of information) affect	Control group had lower CL than "seductive details" group, and performed better at

							workload and skill acquisition and transfer during simulated laparoscopic suturing	immediate testing and transfer
Guru 2015[40]	United States	Quantitative, non-experimental	2 surgical residents, 3 surgical fellows, 5 surgical attendings	Robotic surgical tasks, Simulation	Objective observations or tests	EEG	To compare performance and CL when performing simulated robotic surgical tasks across groups with different levels of experience	Performance was positively associated and CL negatively associated with experience
Haji 2015[41]	Canada	Quantitative, non-experimental	8 1st- and 2nd-year medical students, 5 neurosurgical residents and attendings	Surgical knot tying, Simulation	Objective observations or tests	Paas' Subjective Rating of Mental Effort single item scale, Visual secondary task	To compare CL as measured by secondary task performance and Paas'' single item during simulated surgical knot tying, among subjects of varying expertise	Experts performed better than novices; Experts had lower CL in single- but not dual-task condition; Moving from single- to dual-task increased CL to a greater degree among novices compared with experts; CL declined with practice but eventually plateaued
Haji 2015[42]	Canada	Quantitative, experimental	28 1st-year medical students	Surgical knot tying, Simulation	Quantitative surveys, Objective observations or tests	Paas' Subjective Rating of Mental Effort single item scale, Vibrotactile secondary task	To compare sensitivity of Subjective Rating of Mental Effort and reaction time to a vibrotactile stimulus when varying intrinsic load (simple versus complex task) during simulated surgical knot tying	CL did not differ between simple and complex groups; Subjectively rated CL decreased faster in simple group but no differences in rates of decrease in response time to vibrotactile stimulus
Haji 2016[43]	Canada	Quantitative, experimental	38 2nd-year medical students	Lumbar puncture, Simulation	Quantitative surveys, Objective	Paas' Subjective Rating of	To compare effects of simple versus complex lumbar puncture	Simple group had lower CL and better performance during

					observations or tests	Mental Effort single item scale, Vibrotactile secondary task	training scenario on CL and performance at acquisition, retention and transfer to a “very complex” lumbar puncture scenario	acquisition and retention phases; Simple group had slightly better performance at transfer to very complex scenario (fewer sterility breaches), but no differences in CL at transfer
Hautz 2017[44]	Germany	Quantitative, experimental	49 3rd- and 4th-year medical students	Breast examination, Simulation	Quantitative surveys, Objective observations or tests	CL not measured	To investigate how shame affects CL and learning during simulated breast examination by comparing two groups: high-fidelity simulated breast examination and actual breast examination of a standardized patient; Inferred that shame may impact CL	Students in the standardized patient group had higher shame but also performed better, inferring that shame may not impact CL
Hsieh 2012[45]	Taiwan	Quantitative, experimental	32 graduate students studying nuclear power plant operations	Abnormal operating condition management, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether a support system to help nuclear power plant operators manage abnormal operating procedures would reduce CL and improve performance compared with control condition during simulation	With support system, subjects made decisions more quickly (under both simple and complex conditions) and accurately (under complex conditions only), and CL was lower
Hsu 2015[46]	Taiwan, United Kingdom	Quantitative, non-experimental	18 in-service F16 pilots	Flight, Simulation	Objective observations or tests	Eye tracking	To determine whether CL was associated with performance in simulated flight	Pupil size differed during different parts of the simulation; Better performance associated

								with greater gaze fixation; Experience alone was not associated with CL
Hu 2015[47]	Singapore	Quantitative, non-experimental	47 medical students	Laparoscopic surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine impact of experience level on performance and CL before and after training on simulated laparoscopic tasks, and to determine whether NASA-TLX could be used to assess CL during laparoscopic surgical training	CL decreased and performance improved after training; CL was higher with more complex tasks
Hussein 2016[48]	Egypt, United States	Quantitative, non-experimental	Surgical trainee and mentor	Robotic assisted urologic surgery, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX, EEG	To assess for correlations between CL of a surgical mentor and surgical trainee during robotic-assisted urologic surgery	No correlation between mentor's EEG and trainee's NASA-TLX score; Several correlations were identified between NASA-TLX subscores of the mentor and the trainee, most notably that when the trainee had higher CL, the mentor was paying more attention and was less distracted
Jayaprakash 2016[49]	United States	Quantitative, non-experimental	11 fellow and 3 attending critical care physicians	Assessment of critically ill patient, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare cognitive load of collecting a clinical history for a critically ill patient following, or in parallel with, the primary survey	No difference in CL for series versus parallel timing for collecting clinical history; CL decreased with subsequent cases

Jutric 2017[50]	United States	Quantitative, experimental	5 high school students, 5 attending hepatobiliary surgeons	Targeting tumors for laparoscopic ablation, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL and performance during four different approaches using a novel 3D guidance system for targeting tumors in simulated laparoscopic ablation: in line with and without 3D guidance, and off-axis with and without 3D guidance	Performance was better and CL lower with 3D guidance; Novices initially performed more poorly but with final trial there was no difference between novices and experts
Kataoka 2011[51]	Japan	Quantitative, experimental	32 10 nursing students, 9 "inexperienced" nurses, 13 "experienced" nurses	Operating medication infusion pump, Simulation	Objective observations or tests	NASA-TLX, Eye tracking, Heart rate variability, Respiratory frequency	To examine how time pressure and dual tasking impacted CL of nursing students and nurses during simulated operation of a medication infusion pump	CL as measured by NASA-TLX was higher under both time pressure and dual tasking for all subjects; Eye fixation was lower for all subjects under time pressure; LF/HF was higher for students only under dual tasking
Klein 2009[52]	United States	Quantitative, experimental	15 medical students	Robotic assisted surgery, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Multiple Resources Questionnaire	To compare CL and performance under 2D and 3D vision when performing simulated robotic assisted surgical tasks	In 3D condition (compared with 2D), performance somewhat better (more transfers) and CL lower
Klein 2012[53]	United States	Quantitative, experimental	15 1st-year medical students	Laparoscopic and robotic-assisted surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	Multiple Resources Questionnaire, Dundee Stress State Questionnaire	To compare CL among novices using the Foundations of Laparoscopic Surgery for performing simulated surgical tasks	CL did not differ between systems; Minor differences in performance
Lee 2014[54]	United States	Quantitative, experimental	4 surgical residents, 6 laparoscopic surgery attendings,	Laparoscopic and robotic-assisted	Quantitative surveys, Objective	NASA-TLX	To compare CL and performance among laparoscopic and robotic	Novices and robotic experts had lower CL and better performance

			3 robotic surgery attendings	surgical tasks, Simulation	observations or tests		surgery novices and experts performing simulated laparoscopic and robotic surgery	in robotic versus laparoscopic approach; Laparoscopic experts had lower CL with laparoscopic versus robotic
Lee 2017[55]	United States	Quantitative, non-experimental	6 neuroradiology fellows, 12 attending neuroradiologists	Interpreting neuroradiology images, Workplace	Quantitative surveys	NASA-TLX	To determine whether redesign of a radiology workspace would reduce CL required to completed daily work and improve overall satisfaction	After redesign participants reported reduced levels of CL, better work quality, and greater satisfaction
Lee 2017[56]	United States	Quantitative, experimental	32 surgical residents and fellows	Surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether feedback from an expert mentor during learning of simulated surgical tasks would impact CL and performance, compared with those who completed the learning in a self-directed fashion	Subjects that received feedback performed better and had lower CL than those not receiving feedback
Liang 2010[57]	Taiwan, United States	Quantitative, experimental	22 “junior expert” and 20 “senior expert” aircraft technicians	Aircraft maintenance, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether work instructions administered to aircraft mechanics via an online maintenance assistance platform would be associated with better performance and lower CL compared with simple paper instructions	CL was lower and performance better with online maintenance assistance platform than with paper instructions
Lowndes 2015[58]	United States	Quantitative, experimental	12 medical students	Laparoscopic single site	Quantitative surveys,	NASA-TLX	To compare CL in simulated conventional	Extracorporeal crossing of hands had higher CL

				surgery, Simulation	Objective observations or tests		laparoscopy with two approaches to laparoscopic single site surgery (extracorporeal crossing of hands versus intracorporeal crossing of instruments)	and lower performance than conventional laparoscopy or intracorporeal crossing of instruments
Maxwell 2017[59]	United States	Quantitative, experimental	573 Army National Guard soldiers	Infantry soldier training, Simulation	Surveys (open-ended questions), Objective observations or tests	NASA-TLX	To determine whether virtual, live, or combined virtual and live training of infantry soldiers would be associated with different CL and performance	Combined virtual and live training was associated with lower CL compared with virtual training alone; Combined virtual and live training was associated with better performance than either virtual or live training alone
Mazur 2014[60]	United States	Quantitative, experimental	5 radiation oncology residents, 4 radiation oncology attendings	Radiation therapy planning, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine how CL affects errors during simulated radiation therapy planning	Higher CL was associated with more frequent errors; Residents made errors more often than faculty
Modi 2016[61]	UK	Quantitative, non-experimental	27 "higher surgical trainees"	Laparoscopic suturing, Simulation	Quantitative surveys, Objective observations or tests	SURG-TLX, Heart rate	To determine impact of temporal stressor on resident surgeons' CL and technical performance during simulated laparoscopic suturing	Time pressure increased CL and decreased performance
Mohamed 2014[62]	Canada	Quantitative, non-experimental	8 1st-year gastroenterology residents, 10 practicing gastroenterologists	Endoscopic procedures, Workplace	Quantitative surveys	Modified NASA-TLX	To modify the NASA-TLX for use during endoscopy training and to use the modified tool to map trainees' workload during early	Gastroenterology residents had higher CL than attendings; CL decreased more quickly for upper endoscopy

			and colorectal surgeons				phases of endoscopy training	than colonoscopy procedures
Montero 2011[63]	United States	Quantitative, experimental	14 surgical residents and fellows	Single- and multi-incision laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL and performance associated with simulated standard laparoscopic approach versus single incision laparoscopic surgery approach	Single incision approach was associated with higher CL and lower performance
Moore 2015[64]	China, New Zealand, United Kingdom	Quantitative, non-experimental	32 “qualified and trainee surgeons”	Robotic and laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	Rating Scale for Mental Effort, SURG-TLX, Heart rate variability	To CL and performance among “qualified and trainee” surgeons performing simulated laparoscopic versus robotic surgical tasks	CL was lower and performance higher on robotic versus laparoscopic platform
Moos 2014[65]	United States	Mixed methods, non-experimental	26 pre-service teachers	Teaching, Workplace	Quantitative surveys, Interviews	Paas’ Subjective Rating of Mental Effort single item scale, Investigator-developed multi-item instrument	To characterize how types and levels of CL change throughout teacher training	Overall CL decreased over the course of training; Subjects became increasingly aware of their CL throughout training and felt that their CL as teachers limited their ability to make real-time modifications during teaching
Morris 2006[66]	Australia, Hong Kong	Quantitative, experimental	42 trainee pilots	Flight, Simulation	Objective observations or tests	CL not measured	To determine whether errors occurred more frequently with higher-CL flight simulations	Higher complexity and CL was associated with increased likelihood of errors
Mouraviev 2016[67]	United States	Quantitative, experimental	21 surgical residents	Nephrectomy, Simulation	Quantitative surveys	NASA-TLX	To determine whether simulated nephrectomy using a virtual simulation training environment, or a porcine model, was	CL was similar for both platforms

							associated with differing levels of CL	
Mulcock 2017[68]	United States	Quantitative, non-experimental	140 nursing students	Care of inpatients, Workplace	Quantitative surveys	Investigator-developed single item	To determine whether having continuity of instructors during inpatient clinical rotations was associated with lower levels of stress than conventional placements	Working with the same instructor and in the same facility over first two clinical rotations was associated with reduced stress, which the authors suggest relates to reduce extraneous load
Murai 2010[69]	Japan	Quantitative, non-experimental	10 ship piloting cadets	Ship piloting, Simulation	Objective observations or tests	Heart rate variability	To evaluate the CL of a ship cadet undergoing simulator training	Various aspects of ship navigation task were associated with varying levels of CL
Muresan 2010[70]	United States	Quantitative, experimental	41 medical students	Intracorporeal suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare CL and performance among medical students learning to perform simulated intracorporeal suturing using one of four platforms: laparoscopic suturing, laparoscopic drills, open suturing, and virtual reality drills	Laparoscopic groups had slightly reduced CL compared with other two groups; Performance varied among the four groups but was not clearly linked to level of CL
Naismith 2015[71]	Canada	Mixed methods, non-experimental	38 medical residents	Chest tube insertion, Simulation	Quantitative surveys, Interviews	Paas' Subjective Rating of Mental Effort single item scale, NASA-TLX, Investigator-developed	To compare several measures of CL among medical residents performing simulated chest tube insertion	Total CL scaled to 0-1 scale differed across the different measurement techniques but all were correlated; Qualitative themes included those related to intrinsic load (prior experience, appropriateness for level of training, task

						multi-item instrument		complexity), extraneous load (fidelity, anxiety, impact of hidden curriculum), and germane load (promoting effortful processing, opportunities to observe variations in practice, opportunities for repetitive practice, opportunities for expert feedback)
Rojas 2014[72]	Canada	Quantitative, non-experimental	13 1st-year medical students	One-handed surgical knot tying, Simulation	Objective observations or tests	Vibrotactile secondary task	To determine whether simple reaction time or recognition reaction time as a secondary task measure would be more sensitive to detect differences in CL as medical students learn to tie one-handed surgical knots in a simulated setting	Simple reaction time (detecting a vibrotactile stimulus) was more sensitive than recognition reaction time (detecting only one of two different types of vibrotactile stimuli)
Saleem 2007[73]	United States	Mixed methods, experimental	16 practicing nurses	Using clinical reminders, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To determine whether a redesigned clinical reminder system would be associated with lower CL and improved learnability, usability, and efficiency than the existing system	Redesigned system associated with lower levels of mental demand and frustration (though only statistically significant with one-tailed t-tests); Redesigned system associated with better learnability, usability, and efficiency

Sato 2016[74]	Japan	Quantitative, non-experimental	13 anesthesiology residents, 9 anesthesiology attendings	Administering anesthesia, Simulation	Objective observations or tests	Secondary mathematics task	To determine whether anesthesiology attendings or residents have different CL and performance during simple and complex simulated anesthesia scenarios	No difference in CL or performance for simple scenarios; For complex scenarios residents had higher CL and lower performance than residents; CL was higher in both groups for complex versus simple tasks
Saus 2006[75]	Norway	Quantitative, experimental	40 1st-year police academy students	Police firearm use, Simulation	Quantitative surveys, Objective observations or tests	Heart rate variability	To determine whether situational awareness training would reduce CL and improve performance among police academy students undergoing simulated firearm use training, compared with control group	Situational awareness trained group had lower CL but did not perform differently from control group
Saus 2010[76]	Norway	Quantitative, non-experimental	32 Navy officer cadets	Ship navigation, Simulation	Quantitative surveys, Objective observations or tests	Heart rate variability	To determine whether a low- or high-CL training condition during simulated ship navigation training would impact contributors to learning	In low CL condition learning was impacted by experience, perceived realism and situational awareness; In high CL condition perceived realism and situational awareness impacted learning
Scerbo 2013[77]	United States	Quantitative, non-experimental	7 “expert surgeons,” 11 “intermediate surgeons,” 17 surgical assistants	Laparoscopic surgical tasks, Simulation	Objective observations or tests	Visual secondary task	To determine whether a visual secondary task could distinguish among levels of expertise in simulated laparoscopic tasks	Advancing expertise associated with lower CL, better performance, and faster completion times

Scerbo 2017[78]	United States	Quantitative, non-experimental	12 surgical residents and surgical assistants	Single- and multi-incision laparoscopic surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	Visual secondary task	To examine differences in CL and performance associated with single-incision laparoscopic procedures compared with traditional laparoscopic procedures	CL was higher and performance lower in single-incision versus traditional laparoscopic approach
Scerbo 2017[79]	United States	Quantitative, experimental	22 surgical assistant students, 5 premedical students	Intracorporeal suturing and knot tying, Simulation	Quantitative surveys, Objective observations or tests	Visual secondary task	To assess CL and performance associated with simulated intracorporeal suturing and knot tying over varying levels of transfer	Performance declined on initial reassessment but then returned to baseline; CL was inversely correlated with performance
Seidelman 2010[80]	United States	Quantitative, non-experimental	14 medical students	Laparoscopic ring transfer, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To assess CL and performance associated with single versus tiled surgical displays for a simulated laparoscopic task	Single versus tiled display did not affect CL or accuracy
Sewell 2016[81]	Netherlands, United States	Quantitative, non-experimental	110 1st-year, 179 2nd-year, 174 3rd-year, 12 4th-year gastroenterology fellows	Colonoscopy, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To develop and collect validity for evidence for an instrument to estimate intrinsic, extraneous, and germane load during colonoscopy learning	Intrinsic, extraneous, and germane load were all inversely associated with prior experience
Sewell 2017[82]	Netherlands, United States	Quantitative, non-experimental	110 1st-year, 179 2nd-year, 174 3rd-year, 12 4th-year gastroenterology fellows	Colonoscopy, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To identify characteristics of learners, patients/tasks, environments, and teachers that are associated with intrinsic, extraneous, and germane load during colonoscopy learning	Intrinsic load associated with year in training, prior colonoscopy experience, fatigue, patient tolerance, number of maneuvers performed, and supervisor takeover of colonoscopy; Extraneous load associated with

								fatigue, colonoscopy queue order, supervisor engagement and confidence, and supervisor takeover of colonoscopy; Germane load associated with supervisor engagement
Sexton 2017[83]	Egypt, United States	Quantitative, non-experimental	3 lead surgeons, 3 assistant surgeons, 3 physician assistants, 7 scrub nurses, 11 circulating nurses	Robot-assisted prostatectomy, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX	To investigate how requests during robotic-assisted prostatectomy impact CL and efficiency in the operating room	Number of requests and request duration increased surgeons' CL
Shachak 2009[84]	Canada, Israel	Qualitative	5 residents, 20 practicing primary care physicians	Electronic health record use, Workplace	Interviews	CL not measured, but was estimated qualitatively	To examine the how use of an electronic medical record impacts physicians' CL	Highest CL was perceived with diagnosing, reasoning and treating severe or multiple medical conditions; the electronic medical record was felt to reduce CL, particularly in terms of extraneous load, as results and histories were easy to find
Shewokis 2017[85]	United States	Quantitative, experimental	10 3rd-year medical students	Laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	Functional near infrared spectroscopy	To compare CL and performance in blocked versus random practice during simulated laparoscopic tasks	Compared with blocked practice, random practice was associated with lower CL and better performance
Sibbald 2013[86]	Canada, Netherlands	Quantitative, experimental	191 internal medicine residents	Cardiac auscultation, Simulation	Objective observations or tests	Paas' Subjective Rating of Mental Effort	To determine whether using a checklist during simulated cardiac auscultation learning	Use of checklist was associated with reduced CL and improved

						single item scale	impacts CL and diagnostic accuracy	accuracy compared with no checklist
Sibbald 2013[87]	Canada, Netherlands	Quantitative, non-experimental	15 cardiology fellows	Electrocardiogram interpretation, Simulation	Quantitative surveys, Objective observations or tests	Paas' Subjective Rating of Mental Effort single item scale	To determine whether use of a checklist during experts' verification of electrocardiogram interpretation would reduce errors or impact CL	Using the checklist during verification of interpretation was associated with correction of errors; CL was not affected by use of checklist (i.e., expertise reversal effect was not seen)
Stefanidis 2010[88]	United States	Quantitative, non-experimental	34 medical students	Laparoscopic and robotic intracorporeal suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To compare laparoscopic versus robotic approach to simulated intracorporeal suturing	CL was lower and performance better on robotic versus laparoscopic platform
Strang 2014[89]	Australia, United States	Quantitative, non-experimental	5 Royal Australian Air Force Air Battle Management Officers	Flight, Simulation	Quantitative surveys, Objective observations or tests	Overall Workload Scale, Heart rate variability	To assess for correlations between a subjective measure of CL and heart rate variability among pilots performing flight simulation	Heart rate variability and subjectively rated CL correlated, but the correlations varied across different simulator tasks
Svensson 1993[90]	Sweden	Quantitative, non-experimental	Experiment 1: 21 ground attack pilots Experiment 2: 6 ground attack pilots	Military flight, Simulation	Quantitative surveys	Urinary adrenaline and noradrenalin	To identify contributors to CL during simulated military flight and to correlated CL with performance	Three primary constructs contributed to CL: challenge, effort, and performance; Performance was inversely correlated with CL and challenge positively correlated; CL decreased with training
Szulewski 2014[91]	Canada	Mixed methods, non-experimental	1 medical student, 3 emergency medicine residents, 1 emergency medicine attending	Resuscitation team leading, Simulation	Interviews, Video and audio recordings	CL not measured	To assess feasibility of using eye tracking technology to monitor CL in simulated	Eye tracking device did not interfere with simulation tasks

							resuscitation team leading	
Tattersall 1995[92]	UK	Quantitative, non-experimental	11 trainee flight engineers	Diagnosing flight faults and incidents, Simulation	Quantitative surveys, Objective observations or tests	Subjective Workload Assessment Technique, Heart rate variability	To monitor "naturally occurring" fluctuations in CL during simulated diagnosis of flight faults and incidents, and to compare subjective ratings of CL with heart rate variability	Heart rate variability lower during more challenging simulation tasks; Subjective assessment of CL correlated with heart rate variability
Teel 2009[93]	Japan, United States	Quantitative, non-experimental	10 Marine cadets	Ship navigation, Simulation	Objective observations or tests	Heart rate variability	To evaluate the CL of a cadet undergoing simulator training using heart rate variability	Heart rate variability varied with different tasks during simulation
Theodoraki 2015[94]	Germany	Quantitative, experimental	8 otorhinolaryngology residents, 2 otorhinolaryngology attendings	Functional endoscopic sinus surgery, Workplace	Objective observations or tests	Heart rate variability, Respiratory frequency, Masticator muscle EMG	To compare CL and mental distress of surgeons performing functional endoscopic sinus surgery with and without a navigation system	Navigation system did not impact CL; CL was higher in less experienced group
Thomas 2017[95]	United States	Quantitative, non-experimental	79 practicing nurses	Administering medications to patients, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine how interruptions and distractions impact CL among nurses administering medications to patients in the workplace, and to assess for impact on medication administration errors	CL increased with number of distractions and interruptions; CL did not impact medication administration errors (though the absolute number of errors was very low overall)
Tien 2015[96]	UK	Quantitative, non-experimental	7 "expert" surgeons (attendings and senior residents), 6 "junior" surgeons	Open inguinal hernia repair, Workplace	Quantitative surveys, Objective observations or tests	NASA-TLX, Eye tracking	To assess differences in CL and gaze behavior between expert and junior surgeons during	CL was higher for junior surgeons; Experts had higher fixation frequency and dwell

							open inguinal hernia repair	time during particular parts of the procedure
Tomasko 2012[97]	United States	Quantitative, experimental	31 3rd- and 4th year medical students	Surgical tasks, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Visual secondary task	To compare CL and performance during simulated surgical tasks in sleep-deprived versus non-sleep-deprived states	CL was higher in sleep-deprived group, but performance did not differ
Tremblay 2017[98]	Canada, Netherlands	Mixed methods, non-experimental	143 2nd-year undergraduate pharmacy students	Medication dispensing and patient counseling, Simulation	Quantitative surveys, Objective observations or tests, Focus groups	Investigator-developed multi-item instrument	To determine whether a more authentic "simulated clinical environment" is associated with different intrinsic and extraneous CL compared with a less authentic environment among undergraduate pharmacy students performing simulated medication dispensing and counseling with a standardized patient	Both intrinsic and extraneous load were higher in simulated clinical immersion versus standard environment; Simulated clinical environment was associated with higher self-perceived learning; Students felt they could learn clinical reasoning better in standard environment but could learn to accomplish technical tasks better in simulated clinical environment
Walker 2015[99]	Canada	Qualitative	3 2nd-year dental students, 1 dental attending	Tooth mold production, Simulation	Interviews, Video recordings	Investigator-developed single item	To identify aspects of simulated tooth mold production that were associated with CL, comparing a dental attending with dental students	Many parts of the simulation were noted by students to have high CL, but only a single part was noted by the attending to have high CL
Weigl 2015[100]	Germany	Quantitative, non-experimental	63 attending surgeons	General and orthopedic surgical	Quantitative surveys, Objective	NASA-TLX	To determine whether interruptions during actual surgical operations were	Situational stress was higher during teaching than non-teaching cases; CL was higher for longer

				operations, Workplace	observations or tests		associated with higher CL, and to identify other factors impacting CL, among attending surgeons	cases; Interruptions were associated with increased CL (distraction component of NASA-TLX)
Weigl 2016[101]	Germany	Quantitative, experimental	19 1st- and 2nd-year surgical residents	Vertebroplasty surgery, Simulation	Quantitative surveys, Objective observations or tests	SURG-TLX	To examine the impact of interruptions on resident surgeons' CL and performance on simulated vertebroplasty	CL was higher and performance lower in disrupted scenario; Patient discomfort as source of disruption caused greater increase to CL than telephone call
Weinger 1994[102]	United States	Quantitative, non-experimental	11 1st-year anesthesiology residents, 3 3rd-year anesthesiology residents, 8 certified registered nurse anesthetists	Administering anesthesia, Workplace	Quantitative surveys, Objective observations or tests	Visual secondary task, Borg's workload scale	To develop and to validate objective techniques of CL measurement of anesthesiologists in an actual operating room environment, and to compare CL of novices and experienced anesthesiologists	CL was higher in novices than experienced subjects; Subjective instrument correlated with visual secondary task
Weinger 2000[103]	United States	Quantitative, non-experimental	33 anesthesiology residents, 10 certified registered nurse anesthetists, 83 attending anesthesiologists	Airway management, Workplace	Quantitative surveys	Investigator-developed multi-item instrument	To measure the CL associated with 10 different airway management tasks among three levels of anesthesia providers	CL varied across the 10 tasks; CL did not differ by type or level of provider
Wilson 2011[104]	China, United Kingdom	Quantitative, non-experimental	30 medical students	Laparoscopic tasks, Simulation	Quantitative surveys, Objective observations or tests	SURG-TLX	To develop and validate the SURG-TLX as a measure of CL among surgeons	The results supported discriminant sensitivity of the SURG-TLX to different sources of stress
Workman 2007[105]	United States	Quantitative, experimental	42 3rd- and 4th-year medical students, 3	Interpreting data in intensive care unit, Simulation	Objective observations or tests	Visual secondary task	To determine whether a novel graphical format for information display	CL was lower among medical students using KEGS versus traditional

			board-certified intensivists				(knowledge-enhanced graphical symbols, KEGS) would reduce CL and improve performance interpreting data in simulated critical care medicine setting, as compared with traditional textual display	display; among attendings, CL was lower and accuracy higher using KEGS
Wucherer 2015[106]	Canada, Germany	Quantitative, non-experimental	19 "junior surgeons"	Vertebroplasty surgery, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To estimate CL among "junior surgeons" and assess for relationships with performance, during simulated surgical crisis setting compared with control simulation setting	CL was higher and work slower during crisis compared with control simulation
Young 2010[107]	United States	Quantitative, experimental	16 hypothetical psychiatry residents	Ambulatory psychiatric care, Workplace	Objective observations or tests	CL not measured	To compare predicted ambulatory caseloads for incoming psychiatry residents a traditional model versus novel CL-based model	CL-based model resulted in more evenly balanced patient panels with less inter-caseload variation
Young 2016[108]	Netherlands, United States	Quantitative, experimental	23 2nd-year and 29 6th-year medical students	Patient handovers, Simulation	Quantitative surveys, Objective observations or tests	Paas' Subjective Rating of Mental Effort single item scale	To examine impact of simulated handover complexity on medical students' CL and handover accuracy	CL was negatively correlated with handover accuracy and illness script maturity (measure of experience); Illness script maturity was positively correlated with handover accuracy
Young, 2016[109]	Netherlands, United States	Quantitative, non-experimental	54 2nd-year and 33 3rd-year medical students	Patient handovers, Simulation	Quantitative surveys, Objective	Paas' Subjective Rating of	To develop an instrument to measure subtypes of CL among	2-factor model for CL developed including intrinsic and germane

					observations or tests	Mental Effort single item scale, Investigator-developed multi-item instrument	medical students performing simulated handovers, and to assess for relationships between intrinsic and germane load with experience level and performance	load; intrinsic load and germane load were higher in less experienced subjects; No significant associations between CL and performance; Paas' single item positively correlated with intrinsic but not germane load
Young, 2016[110]	Netherlands, United States	Quantitative, non-experimental	23 2nd-year and 29 6th-year medical students	Patient handovers, Simulation	Quantitative surveys	Paas' Subjective Rating of Mental Effort single item scale, Investigator-developed multi-item instrument	To provide evidence for validity of a revised instrument to measure CL during handovers, and to assess for relationships between CL, case complexity, and experience	Intrinsic and germane load formed a single factor, which was positively correlated with handover complexity and negatively correlated with experience
Yu, 2016[111]	United States	Quantitative, non-experimental	4 anesthesiologists, 12 certified registered nurse anesthetists, 38 circulating nurses, 35 surgical technicians, 26 surgical assistants, 45 surgical residents, 32 attending surgeons	Surgical operations, Workplace	Quantitative surveys, Objective observations or tests	SURG-TLX	To quantify and compare CL among surgical team members across different surgical techniques and specialties	CL was highest among surgical residents (most novice team members); Ancillary staff (circulating nurse, surgical technician) had lowest CL; Surgical duration positively associated with overall team CL
Yurko 2010[112]	United States	Quantitative, non-experimental	28 2nd-year medical students and senior premedical students	Laparoscopic suturing, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine relationships between CL and performance among medical and premedical students performing	CL and performance were negatively correlated; Higher CL associated with more damage to porcine

							simulated laparoscopic suturing, first on Fundamentals of Laparoscopic Surgery model, and then on transfer to porcine model	tissues; Performance increased and CL decreased on post-training testing, but then reversed on transfer to a porcine model
Zheng 2010[113]	Canada	Quantitative, non-experimental	12 surgical residents, 9 surgical fellows and attendings	Laparoscopic tasks, Simulation	Objective observations or tests	Visual secondary task	To compare CL and performance while performing simulated laparoscopic tasks among experienced versus novice laparoscopic surgeons	Experienced surgeons had lower CL and better performance than novice surgeons
Zheng 2012[114]	Canada	Quantitative, experimental	5 novice and 5 expert surgeons	Laparoscopic and natural orifice transluminal endoscopic surgery tasks, Simulation	Objective observations or tests	NASA-TLX, Visual secondary task	To compare CL and performance among novice and expert surgeons performing simulated laparoscopic and natural orifice transluminal endoscopic surgery (NOTES)	CL was higher in NOTES versus laparoscopic procedures (based on secondary task, but not based on NASA-TLX) but this was attenuated by prior experience; Performance lower in NOTES procedures and higher among those with greater prior experience
Zheng 2012[115]	Canada	Quantitative, non-experimental	23 surgical residents, fellows, and attendings	Laparoscopic cholecystectomy, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX, Eye tracking	To correlate eye tracking and NASA-TLX as means to estimate CL during simulated laparoscopic cholecystectomy	Higher CL was associated with fewer blinks and shorter blink duration; Performance did not vary with eye tracking data
Zheng 2014[116]	Canada	Quantitative, experimental	12 surgical residents, 3 laparoscopic fellows, 2 surgical attendings	Using surgical instruments, Simulation	Quantitative surveys, Objective observations or tests	NASA-TLX	To examine how complexity (open versus laparoscopic versus endoscopic approach) impacts surgeons' CL	CL was higher in most complex approach (endoscopic) versus least complex approach (open); Time to

							and performance during simulated use of surgical tools	complete task higher in more complex approaches; With practice, performance improved but CL did not change

Abbreviations: CL, cognitive load; EEG, electroencephalogram; EMG, electromyography; LF/HF, low frequency/high frequency (higher ratio suggests higher sympathetic tone and higher CL); NASA-TLX, NASA Task Load Index; SURG-TLX, Surgery Task Load Index

[a]Description of participants is provided using the level of detail included in each study manuscript.

[b]For consistency and ease of table interpretation, we use the term CL (cognitive load) when referring to the constructs of cognitive load, mental effort, and mental workload.

REFERENCES FOR SUPPLEMENTAL TABLE B

1. Andersen SA, Mikkelsen PT, Konge L, Caye-Thomasen P, Sorensen MS. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *Laryngoscope.* 2015;126:E74-E79.
2. Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen M. Cognitive Load in Mastoidectomy Skills Training: Virtual Reality Simulation and Traditional Dissection Compared. *J Surg Educ.* 2016;73(1):45-50.
3. Andersen SA, Konge L, Cayé-Thomasen P, Sørensen MS. Retention of mastoidectomy skills after virtual reality simulation training. *JAMA Otolaryngol Head Neck Surg.* 2016;142(7):635-640.
4. Andrade AD, Cifuentes P, Mintzer MJ, Roos BA, Anam R, Ruiz JG. Simulating Geriatric Home Safety Assessments in a Three-Dimensional Virtual World. *Gerontol Geriatr Educ.* 2012;33(3):233-252.
5. Arico P, Borghini G, Di Flumeri G, et al. Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Front Hum Neurosci.* 2016;10:539.
6. Attrill S, Lincoln M, McAllister S. Supervising international students in clinical placements: perceptions of experiences and factors influencing competency development. *BMC Med Educ.* 2016;16:180.
7. Avansino J, Leu MG. Effects of CPOE on provider cognitive workload: A randomized crossover trial. *Pediatrics.* 2012;130(3):e547-e552.
8. Bertram DA, Hershey CO, Opila DA, Quirin O. A measure of physician mental work load in internal medicine ambulatory care clinics. *Med Care.* 1990;28(5):458-467.
9. Bertram DA, Opila DA, Brown JL, et al. Measuring physician mental workload: reliability and validity assessment of a brief instrument. *Med Care.* 1992;30(2):95-104.
10. Bharathan R, Vali S, Setchell T, Miskry T, Darzi A, Aggarwal R. Psychomotor skills and cognitive load training on a virtual reality laparoscopic simulator for tubal surgery is effective. *Eur J Obstet Gynecol Reprod Biol.* 2013;169(2):347-352.
11. Blissett S, Morrison D, McCarty D, Sibbald M. Should learners reason one step at a time? A randomised trial of two diagnostic scheme designs. *Med Educ.* 2017;51(4):432-441.
12. Boet S, Sharma B, Pigford AA, Hladkowicz E, Rittenhouse N, Grantcharov T. Debriefing decreases mental workload in surgical crisis: A randomized controlled trial. *Surgery.* 2017;161(5):1215-1220.
13. Britt RC, Scerbo MW, Montano M, Kennedy RA, Prytz E, Stefanidis D. Intracorporeal suturing: Transfer from Fundamentals of Laparoscopic Surgery to cadavers results in substantial increase in mental workload. *Surgery.* 2015;158(5):1428-1433.
14. Broyles TW, Epler CM, Waknine JW. 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.* 2011;36(1):49-66.
15. Byrne AJ, Sellen AJ, Jones JG. Errors on anaesthetic record charts as a measure of anaesthetic performance during simulated critical incidents. *Br J Anaesth.* 1998;80(1):58-62.
16. Byrne AJ, Murphy A, McIntyre O, Tweed N. The relationship between experience and mental workload in anaesthetic practice: an observational study. *Anaesthesia.* 2013;68(12):1266-1272.
17. Cao CGL, Zhou M, Jones DB, Schwaitzberg SD. Can surgeons think and operate with haptics at the same

time? *J Gastrointest Surg.* 2007;11(11):1564-1569.
18. Chen R, Grierson LE, Norman GR. Evaluating the impact of high- and low-fidelity instruction in the development of auscultation skills. *Med Educ.* 2015;49(3):276-285.
19. Chen R, Grierson L, Norman G. Manipulation of cognitive load variables and impact on auscultation test performance. *Adv Health Sci Educ Theory Pract.* 2015;20(4):935-952.
20. Chowriappa A, Raza SJ, Fazili A, et al. Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: A multi-institutional randomised controlled trial. *BJU Int.* 2015;115(2):336-345.
21. Chui MA, Stone JA. Exploring information chaos in community pharmacy handoffs. *Res Social Adm Pharm.* 2014;10(1):195-203.
22. Crosby JJ, Parkinson SR. A dual task investigation of pilots' skill level. *Ergonomics.* 1979;22(12):1301-1313.
23. Cummings ML, Smith CA. Improving precision vertical flight operations through a direct-perception-action display. *Journal of Aircraft.* 2009;46(4):1270-1279.
24. Dahlstrom N, Nahlinder S. Mental workload in aircraft and simulator during basic civil aviation training. *International Journal of Aviation Psychology.* 2009;19(4):309-325.
25. Dankbaar ME, Alsma J, Jansen EE, van Merriënboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ Theory Pract.* 2016;21(3):505-521.
26. Davis DHJ, Oliver M, Byrne AJ. A novel method of measuring the mental workload of anaesthetists during simulated practice. *Br J Anaesth.* 2009;103(5):665-669.
27. de Man FR, Erwteman M, van Groeningen D, et al. The effect of audible alarms on anaesthesiologists' response times to adverse events in a simulated anaesthesia environment: a randomised trial. *Anaesthesia.* 2014;69(6):598-603.
28. Di Stasi LL, Díaz-Piedra C, Ruiz-Rabelo JF, Rieiro H, Sanchez Carrion JM, Catena A. Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: Gaze-based indices. *Appl Ergon.* 2017;65:168-174.
29. Doig AK, Albert RW, Syroid ND, Moon S, Agutter JA. Graphical arterial blood gas visualization tool supports rapid and accurate data interpretation. *Comput Inform Nurs.* 2011;29(4 Suppl):TC53-60.
30. Dominessy ME, Lukas JH, Malkin FJ, Monty RA, Oatman LC. The effect of information display formats on helicopter pilots' target acquisition and flying performance. *Mil Psychol.* 1991;3(3):163-176.
31. Donato AA, Kaliyadan AG, Wasser T. Self-directed study using MP3 players to improve auscultation proficiency of physicians: a randomized, controlled trial. *J Contin Educ Health Prof.* 2014;34(2):131-138.
32. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. 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.* 2012;17(1):65-79.
33. Durning S, Artino AR, Jr., Pangaro L, van der Vleuten CP, Schuwirth L. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ.* 2011;45(9):927-938.
34. Fenik Y, Celebi N, Wagner R, et al. Prepackaged central line kits reduce procedural mistakes during central line insertion: a randomized controlled prospective trial. *BMC Med Educ.* 2013;13:60.
35. Flindall IR, Leff DR, Pucks N, Sugden C, Darzi A. The Preservation of Cued Recall in the Acute Mentally Fatigued

State: A Randomised Crossover Study. *World J Surg.* 2016;40:56-65.
36. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ.* 2012;46(11):1055-1062.
37. Fraser K, Huffman J, Ma I, et al. The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. *Chest.* 2014;145(5):958-963.
38. Gaba DM, Lee T. Measuring the workload of the anesthesiologist. *Anesth Analg.* 1990;71(4):354-361.
39. Gardner AK, Clanton J, Jabbour, II, Scott L, Scott DJ, Russo MA. Impact of seductive details on the acquisition and transfer of laparoscopic suturing skills: Emotionally interesting or cognitively taxing? *Surgery.* 2016;160(3):580-585.
40. Guru KA, Esfahani ET, Raza SJ, et al. Cognitive skills assessment during robot-assisted surgery: separating the wheat from the chaff. *BJU Int.* 2015;115(1):166-174.
41. Haji FA, Khan R, Regehr G, Drake J, de Ribaupierre S, Dubrowski A. Measuring cognitive load during simulation-based psychomotor skills training: sensitivity of secondary-task performance and subjective ratings. *Adv Health Sci Educ Theory Pract.* 2015;20(5):1237-1253.
42. Haji FA, Rojas D, Childs R, Ribaupierre S, Dubrowski A. Measuring cognitive load: performance, mental effort and simulation task complexity. *Med Educ.* 2015;49(8):815-827.
43. Haji FA, Cheung JJ, Woods N, Regehr G, de Ribaupierre S, Dubrowski A. Thrive or overload? The effect of task complexity on novices' simulation-based learning. *Med Educ.* 2016;50(9):955-968.
44. Hautz WE, Schroder T, Dannenberg KA, et al. Shame in Medical Education: A Randomized Study of the Acquisition of Intimate Examination Skills and Its Effect on Subsequent Performance. *Teach Learn Med.* 2017;29(2):196-206.
45. Hsieh MH, Hwang SL, Liu KH, Liang SFM, Chuang CF. A decision support system for identifying abnormal operating procedures in a nuclear power plant. *Nuclear Engineering and Design.* 2012;249:413-418.
46. Hsu CK, Lin SC, Li WC. Visual movement and mental-workload for pilot performance assessment. *Lecture Notes in Computer Science.* 2015;9174:356-364.
47. Hu JS, Lu J, Tan WB, Lomanto D. Training improves laparoscopic tasks performance and decreases operator workload. *Surg Endosc.* 2015;30:1742-1746.
48. Hussein AA, Shafiei SB, Sharif M, et al. Technical Mentorship during Robot-Assisted Surgery: A Cognitive Analysis? *BJU Int.* 2016;118(3):429-436.
49. Jayaprakash N, Ali R, Kashyap R, Bennett C, Kogan A, Gajic O. The incorporation of focused history in checklist for early recognition and treatment of acute illness and injury. *BMC Emerg Med.* 2016;16(1):35.
50. Jutric Z, Grendar J, Brown WL, et al. Novel Simulation Device for Targeting Tumors in Laparoscopic Ablation: A Learning Curve Study. *Surg Innov.* 2017;24(5):492-498.
51. Kataoka J, Sasaki M, Kanda K. Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Jpn J Nurs Sci.* 2011;8(1):47-56.
52. Klein MI, Lio CH, Grant R, Carswell CM, Strup S. A mental workload study on the 2d and 3d viewing conditions of the da Vinci surgical robot. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting.* 2009;2:1186-1190.
53. Klein MI, Warm JS, Riley MA, et al. Mental workload and stress perceived by novice operators in the laparoscopic and robotic minimally

invasive surgical interfaces. *J Endourol.* 2012;26(8):1089-1094.
54. Lee GI, Lee MR, Clanton T, Sutton E, Park AE, Marohn MR. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. *Surg Endosc.* 2014;28(2):456-465.
55. Lee MH, Schemmel AJ, Pooler BD, et al. Radiology Workflow Dynamics: How Workflow Patterns Impact Radiologist Perceptions of Workplace Satisfaction. *Acad Radiol.* 2017;24(4):483-487.
56. Lee GI, Lee MR. 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.* 2018;32(1):62-72.
57. Liang GF, Lin JT, Hwang SL, Wang EM, Patterson P. Preventing human errors in aviation maintenance using an on-line maintenance assistance platform. *International Journal of Industrial Ergonomics.* 2010;40(3):356-367.
58. Lowndes B, Abdelrahman A, McCrory B, Hallbeck S. A Preliminary Study of Novice Workload and Performance During Surgical Simulation Tasks for Conventional vs. Single Incision Laparoscopic Techniques. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting.* 2015:498-502.
59. Maxwell DZ, J. Large Scale Testing and Evaluation of Virtual Environments for Infantry Soldier Tasks Comparing Mental Effort for Live Versus Virtual Training Performance Assessments. *MODSIM World 2017*; April 25-27, 2017; Virginia Beach, VA.
60. Mazur LM, Mosaly PR, Hoyle LM, Jones EL, Chera BS, Marks LB. Relating physician's workload with errors during radiation therapy planning. *Pract Radiat Oncol.* 2014;4(2):71-75.
61. Modi H, Leff D, Singh H, Darzi A. The influence of cognitive load on technical ability among surgical trainees. *Int J Surg.* 2016;36:S34-S35.
62. Mohamed R, Raman M, Anderson J, McLaughlin K, Rostom A, Coderre S. 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.* 2014;28(3):155-159.
63. Montero PN, Acker CE, Heniford BT, Stefanidis D. Single incision laparoscopic surgery (SILS) is associated with poorer performance and increased surgeon workload compared with standard laparoscopy. *Am Surg.* 2011;77(1):73-77.
64. Moore LJ, Wilson MR, McGrath JS, Waine E, Masters RSW, Vine SJ. Surgeons' display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy. *Surg Endosc.* 2015;29(9):2553-2560.
65. Moos DC, Pitton D. Student Teacher Challenges: Using the Cognitive Load Theory as an Explanatory Lens. *Teaching Education.* 2014;25(2):127-141.
66. Morris CH, Leung YK. Pilot mental workload: how well do pilots really perform? *Ergonomics.* 2006;49(15):1581-1596.
67. Mouraviev V, Klein M, Schommer E, et al. Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules. *J Robot Surg.* 2016;10(1):49-56.
68. Mulcock PM, Grassley J, Davis M, White K. Beyond the Dedicated Education Unit: Using Cognitive Load Theory to Guide Clinical Placement. *J Nurs Educ.* 2017;56(2):105-109.
69. Murai K, Ishikura A, Hayashi Y, et al. Toward evaluation of ship bridge teamwork for marine simulator

training based on heart rate variability. *World Automation Congress.* 2010:1-6.
70. Muresan III C, Lee TH, Seagull J, Park AE. Transfer of training in the development of intracorporeal suturing skill in medical student novices: A prospective randomized trial. *Am J Surg.* 2010;200(4):537-541.
71. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ.* 2015;48(8):805-814.
72. Rojas D, Haji F, Shewaga R, Kapralos B, Dubrowski A. The impact of secondary-task type on the sensitivity of reaction-time based measurement of cognitive load for novices learning surgical skills using simulation. *Stud Health Technol Inform.* 2014;196:353-359.
73. Saleem JJ, Patterson ES, Militello L, et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. *J Am Med Inform Assoc.* 2007;14(5):632-640.
74. Sato H, Miyashita T, Kawakami H, Nagamine Y, Takaki S, Goto T. Influence of Mental Workload on the Performance of Anesthesiologists during Induction of General Anesthesia: A Patient Simulator Study. *BioMed Research International.* 2016;2016:1058750.
75. Saus E, Johnsen BH, Eid J, Riisem PK, Andersen R, Thayer JF. The effect of brief situational awareness training in a police shooting simulator: an experimental study. *Mil Psychol.* 2006;18(Suppl):S3-S21.
76. Saus E, Johnsen BH, Eid J. Perceived learning outcome: the relationship between experience, realism and situation awareness during simulator training. *Int Marit Health.* 2010;62(4):258-264.
77. Scerbo MW, Kennedy A, Montano M, Britt RC, Davis SS, Stefanidis D. A Spatial Secondary Task for Measuring Laparoscopic Mental Workload: Differences in Surgical Experience. *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting.* 2013:728-732.
78. Scerbo MW, Britt RC, Stefanidis D. Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. *Am J Surg.* 2017;213(2):244-248.
79. Scerbo MW, Britt RC, Montano M, Kennedy RA, Prytz E, Stefanidis D. Effects of a retention interval and refresher session on intracorporeal suturing and knot tying skill and mental workload. *Surgery.* 2017;161(5):1209-1214.
80. Seidelman W, Carswell CM, Lio CH, et al. Potential performance costs associated with large-format tiled displays for surgical visualization. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting.* 2010;2:1430-1434.
81. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. *Med Educ.* 2016;50(6):682-692.
82. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. 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.* 2017;92(11):1622-1631.
83. Sexton K, Johnson A, Gotsch A, Hussein AA, Cavuoto L, Guru KA. Anticipation, teamwork and cognitive load: chasing efficiency during robot-assisted surgery. *BMJ Qual Saf.* 2018;27(2):148-154.
84. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. *J Gen Intern Med.* 2009;24(3):341-348.

85. Shewokis PA, Shariff FU, Liu Y, Ayaz H, Castellanos A, Lind DS. Acquisition, retention and transfer of simulated laparoscopic tasks using fNIR and a contextual interference paradigm. *Am J Surg.* 2017;213(2):336-345.
86. Sibbald M, de Bruin AB, Cavalcanti RB, van Merriënboer JJG. Do you have to re-examine to reconsider your diagnosis? Checklists and cardiac exam. *BMJ Qual Saf.* 2013;22(4):333-338.
87. Sibbald M, de Bruin ABH, van Merriënboer JJG. Checklists improve experts' diagnostic decisions. *Med Educ.* 2013;47(3):301-308.
88. Stefanidis D, Wang F, Korndorffer JR, Jr., Dunne JB, Scott DJ. Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload. *Surg Endosc.* 2010;24(2):377-382.
89. Strang AJ, Best C, Funke GJ. Heart rate correlates of mental workload in a large-scale air-combat simulation training exercise. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting.* 2014:2325-2329.
90. Svensson E, Angelborg-Thanderz M, Sjöberg L. Mission challenge, mental workload and performance in military aviation. *Aviat Space Environ Med.* 1993;64(11):985-991.
91. Szulewski A, Howes D. Combining first-person video and gaze-tracking in medical simulation: A technical feasibility study. *The Scientific World Journal.* 2014;2014(975752):eCollection 2014.
92. Tattersall AJ, Hockey GRJ. Level of operator control and changes in heart rate variability during simulated flight maintenance. *Hum Factors.* 1995;37(4):682-698.
93. Teel S, Sanders J, Parrott DS, et al. Evaluation of marine simulator training based on heart rate variability. *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics.* 2009:233-238.
94. Theodoraki MN, Ledderose GJ, Becker S, et al. Mental distress and effort to engage an image-guided navigation system in the surgical training of endoscopic sinus surgery: a prospective, randomised clinical trial. *Eur Arch Otorhinolaryngol.* 2015;272(4):905-913.
95. Thomas L, Donohue-Porter P, Stein Fishbein J. Impact of Interruptions, Distractions, and Cognitive Load on Procedure Failures and Medication Administration Errors. *J Nurs Care Qual.* 2017;32(4):309-317.
96. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang GZ, Darzi A. Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surg Endosc.* 2015;29(2):405-413.
97. Tomasko JM, Pauli EM, Kunselman AR, Haluck RS. Sleep deprivation increases cognitive workload during simulated surgical tasks. *Am J Surg.* 2012;203(1):37-43.
98. Tremblay ML, Lafleur A, Leppink J, Dolmans DH. The simulated clinical environment: Cognitive and emotional impact among undergraduates. *Med Teach.* 2017;39(2):181-187.
99. Walker J, von Bergmann H. Lessons from a pilot project in cognitive task analysis: the potential role of intermediates in preclinical teaching in dental education. *J Dent Educ.* 2015;79(3):286-294.
100. Weigl M, Antoniadis S, Chiapponi C, Bruns C, Sevdalis N. The impact of intra-operative interruptions on surgeons' perceived workload: an observational study in elective general and orthopedic surgery. *Surg Endosc.* 2015;29(1):145-153.
101. Weigl M, Stefan P, Abhari K, et al. Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-scale simulated

procedure. *Surg Endosc.* 2016;30(2):559-566.
102. Weinger MB, Herndon OW, Zornow MH, Paulus MP, Gaba DM, Dallen LT. An objective methodology for task analysis and workload assessment in anesthesia providers. *Anesthesiology.* 1994;80(1):77-92.
103. Weinger MB, Vredenburgh AG, Schumann CM, et al. Quantitative description of the workload associated with airway management procedures. *J Clin Anesth.* 2000;12(4):273-282.
104. Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RS. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World J Surg.* 2011;35(9):1961-1969.
105. Workman M, Lesser MF, Kim J. An exploratory study of cognitive load in diagnosing patient conditions. *Int J Qual Health Care.* 2007;19(3):127-133.
106. Wucherer P, Stefan P, Abhari K, et al. Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis. *IEEE Trans Med Imaging.* 2015;34(8):1730-1737.
107. Young JQ, Niehaus B, Lieu SC, O'Sullivan PS. Improving resident education and patient safety: a method to balance initial caseloads at academic year-end transfer. *Acad Med.* 2010;85(9):1418-1424.
108. Young JQ, van Dijk SM, O'Sullivan PS, Custers EJ, Irby DM, ten Cate O. Influence of learner knowledge and case complexity on handover accuracy and cognitive load: results from a simulation study. *Med Educ.* 2016;50(9):969-978.
109. Young JQ, Irby DM, Barilla-LaBarca ML, ten Cate O, O'Sullivan PS. Measuring cognitive load: mixed results from a handover simulation for medical students. *Perspect Med Educ.* 2016;5(1):24-32.
110. Young JQ, Boscardin CK, van Dijk SM, et al. Performance of a cognitive load inventory during simulated handoffs: Evidence for validity. *SAGE Open Med.* 2016;4:2050312116682254.
111. Yu D, Lowndes B, Thiels C, et al. Quantifying Intraoperative Workloads Across the Surgical Team Roles: Room for Better Balance? *World J Surg.* 2016;40(7):1565-1574.
112. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthc.* 2010;5(5):267-271.
113. Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanström LL. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surg Endosc.* 2010;24(1):45-50.
114. Zheng B, Rieder E, Cassera MA, et al. Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. *Surg Endosc.* 2012;26(5):1352-1358.
115. Zheng B, Jiang X, Tien G, Meneghetti A, Panton ONM, Atkins MS. Workload assessment of surgeons: Correlation between NASA TLX and blinks. *Surg Endosc.* 2012;26(10):2746-2750.
116. Zheng B, Fu B, Al-Tayeb TA, Hao YF, Qayumi AK. Mastering instruments before operating on a patient: the role of simulation training in tool use skills. *Surg Innov.* 2014;21(6):637-642.

Chapter 5

Learning challenges, teaching strategies, and cognitive load: experience of seasoned endoscopy teachers

Sewell, JL

Bowen JL

ten Cate, O

O'Sullivan, PS

Shah, B

Boscardin, CK

Accepted for publication as:

Learning challenges, teaching strategies, and cognitive load: experience of seasoned endoscopy teachers

Academic Medicine

This is a non-final version of an article published in final form in *Academic Medicine.*

ABSTRACT

Purpose: Learners of medical procedures must develop, refine and apply schemas for both cognitive and psychomotor constructs. This dual-task condition may strain working memory capacity. Procedures with limitations in visual and tactile information may pose additional risk of cognitive overload. The authors sought to elucidate how experienced procedural teachers perceived learners' challenges and their own teaching strategies in the exemplar setting of gastrointestinal endoscopy.

Methods: The authors interviewed 22 experienced endoscopy teachers in the United States, Canada, and the Netherlands between May 2016 and March 2019 and performed thematic analysis using template analysis method. Cognitive load theory informed data interpretation and analysis.

Results: Participants described taking steps to "diagnose" trainee ability and identify struggling trainees. They described learning challenges related to: trainees (performance over mastery goal orientation, low self-efficacy, lack of awareness), tasks (psychomotor challenges, mental model development, tactile understanding), teachers (teacher-trainee relationship, inadequate teaching, teaching variability), and settings (internal and external distractions, systems issues). Participants described employing teaching strategies that could: match intrinsic load to learners' levels (teaching along developmental continuum, motor instruction, technical assistance/takeover), minimize extraneous load (optimize environment, systems solutions, emotional support, define expectations), and optimize germane load (promote mastery, teach schemas, stop and focus).

Conclusions: Participants' stories provide insight into challenges trainees may experience while learning to perform complex medical procedures with limitations in sensory channels, as well as teaching strategies that may address these challenges at individual and systems levels. Using the framework of cognitive load theory, the authors provide recommendations for procedural teachers.

INTRODUCTION

Learning procedural skills is a challenge across the continuum of medical education. According to cognitive load theory (CLT), learning occurs when cognitive schemas (cognitive representations of constructs that are stored in the long-term memory) are formed and refined.[1,2] During procedural skills training, trainees must develop, refine, and activate complex schemas for both cognitive and psychomotor constructs. This dual task requirement may induce cognitive overload, which interferes with learning.[2]

The primary senses used in medical procedures are sight and touch, yet certain procedures involve substantial limitations in visual and tactile information. For example, to perform imaging-guided biopsy, the interventional radiologist inserts a needle deep into the body of a patient, which exists in three dimensions. However, the radiologist must use limited two-dimensional images to guide needle insertion;[3] the visual information needed to complete the procedure is limited in detail and accessibility. Similarly, robotic surgery provides limited tactile information.[4] The surgeon cannot directly feel the tissue on which she is operating. Lacking haptic feedback, she must infer the feel of the tissue from information that robotic controls provide; her tactile sense is constrained by the technology and equipment she must use to optimize surgical outcomes.

Gastrointestinal endoscopy is another class of procedures with visual and tactile constraints.[5] A gastrointestinal endoscope is a long flexible tube (ranging from approximately 80 to 200 cm) with a camera at the tip that the performing provider pushes into and then pulls out of the patient's body, and controls using dials connected to the tip by cables. During an endoscopic procedure, the instrument's three-dimensional configuration can change dramatically within the patient's gastrointestinal tract,[5] yet the performing provider sees only a two-dimensional, cross-sectional view on the video monitor. Additionally, the equipment transmits less direct tactile feedback[5] than other procedures performed directly with hands on tissues, or with shorter, rigid instruments. These sensory limitations make it more challenging to develop an accurate understanding of what is happening with the instrument inside of the patient. These limitations contribute substantial complexity to endoscopy learning, reflected in the large numbers of endoscopic procedures required to attain competence.[6]

Prior studies of sensory-constrained procedural fields including surgery and gastrointestinal endoscopy have yielded important findings such as deconstructing individual tasks required to develop competence,[7,8] developing and testing methods and instruments to assess technical competence,[6,9-11] and confirming the utility of simulation-based education.[12-15] However, specific learning challenges, and specific teaching strategies to address those challenges, are not well studied. Greater understanding could inform educational innovations to promote learning, which could in turn lead to a better-trained junior procedural workforce and improved patient outcomes. To elucidate the challenges that learners of sensory-constrained procedures may experience, and to characterize how teachers help learners overcome these challenges, we studied the experiences of teachers in the exemplar setting of gastrointestinal endoscopy.

Considering the complexity of procedural learning, CLT provides an apt theoretical framework. CLT is a cognitive learning theory focused on limitations of the working memory.[1,2] When working memory is overloaded, learning and performance are impaired. Three main types of cognitive load impose on working memory. *Intrinsic load* occurs as learners perform essential components of a task. *Extraneous load* occurs when learners direct mental effort to distractions or other nonessential stimuli. *Germane load* occurs when learners engage in cognitive activities that promote formation and refinement of learning schema (in other words, when they are learning). Working memory can manage only a few pieces of information at a time and is easily overwhelmed,[16] so the absolute and relative amount of each cognitive load type is critically important. Extraneous load should be minimized and intrinsic load matched to the learner's prior experience; this provides working memory space for activities that promote germane load and learning.

Despite relevance of CLT to procedural skills training in general,[17] and endoscopy learning in particular,[18-20] little qualitative research has investigated CLT in procedural health professions education (HPE) workplace settings.[17] We know relatively little about the specific challenges that teachers witness procedural learners experience or how CLT might inform teachers' educational approaches. Therefore, the purpose of this study is to identify teachers' perspectives of learning challenges and teaching strategies during gastrointestinal endoscopy training, using CLT as a guiding theoretical framework. We elicited perceptions of experienced teachers of gastrointestinal endoscopy to address three primary research questions:

1) What specific learner challenges do experienced teachers observe during gastrointestinal endoscopy training?
2) What strategies do experienced teachers use to promote learning of gastrointestinal endoscopy?
3) How do these teaching strategies align with CLT?

METHODS

Design: We performed a qualitative study undertaking thematic analysis of interviews with experienced endoscopy teachers. We adopted a "critical realist" epistemology,[21] which couples ontological realism (i.e., there exist identifiable learning challenges and teaching strategies) with epistemological constructivism and relativism (i.e., how teachers perceive and report these will vary).

Participants and procedures: We sought perspectives of experienced endoscopy teachers who had 5 or more years of ongoing endoscopy teaching experience. Compared with trainees, teachers have greater insight into learning challenges and teaching strategies, informed by teaching numerous trainees and performing thousands of procedures. We sought perspectives of experienced, rather than novice, teachers; though both would have content knowledge (understanding of how to perform endoscopy), experienced teachers would have greater pedagogical knowledge[22] (understanding of how to teach endoscopy) necessary to address our objectives. We selected the five-year mark, because by this time, teachers would have sufficient experience to complete the 500 colonoscopies generally considered necessary to develop expertise,[6] and to have worked with diverse learners through diverse endoscopic cases to develop pedagogical expertise. We solicited those who were program directors, had specialized training in education, and held leadership positions in gastrointestinal endoscopy societies. To enhance diversity of perspectives, we recruited participants from multiple specialties that perform endoscopic procedures including adult gastroenterologists, pediatric gastroenterologists and advanced endoscopists (gastroenterologists with additional training who perform specialized endoscopic procedures), as well as colorectal surgeons and family medicine physicians. We sought participants internationally (from the United States, Canada, and the Netherlands) to broaden perspectives. We identified participants using several methods (faculty known to the investigators, society leaders, those who have published on endoscopy training, and snowball sampling) and recruited participants by email.

Participants completed a 20-60 minute semi-structured in-person or telephone interview with JLS, which was audio-recorded, de-identified, and transcribed. Interviews were performed between May 2016 and March 2019. Participants provided demographics including specialty, types of learners, gender, and years of teaching. Recruitment ceased when we reached information sufficiency.[23] This determination was initially made by authors who read and coded transcripts (JLS, JLB, CKB) and was approved by all authors.

Interview protocol: The authors generated the primary interview questions (List 1) through review of the literature and iterative discussions. The first two questions employed a critical incident technique[24] to enhance richness of stories and provide tangible examples of challenges and strategies. Concepts of CLT informed development of the interview guide, but CLT-specific questions were not included. The full interview guide is available as Supplemental Digital Appendix 1.

List 1. Primary interview questions when interviewing 22 experienced endoscopy teachers regarding learner challenges and teaching strategies during gastrointestinal endoscopy training.

1. *Please think about an endoscopic procedure that you thought was particularly challenging for the learner. Can you tell me about that encounter? What was your approach to helping the learner?*
2. *Can you think of another, different, situation in which a learner was struggling with a different type of endoscopic procedure? Can you tell me about that experience? How did the challenges they experienced differ from the first example? How did your teaching approach differ?*
3. *In general, what do you see as the most significant challenges that learners experience during endoscopy training? What strategies can most effectively help learners overcome those challenges?*
4. *In your training program, how do learners learn to perform endoscopic procedures? Are there curricular or systemic factors that affect your learners' endoscopy training in a positive or negative manner?*

Reflexivity: The primary author (JLS) is an academic gastroenterologist with over 7 years' experience teaching gastrointestinal endoscopy, and an educational researcher. Three authors (CKB, OtC, PSO'S) have expertise in education research. Two additional clinicians contributed. BS is an academic gastroenterologist with nearly a decade of endoscopy teaching experience. JLB is an academic general internist. Beyond clinical expertise, BS and JLB provided perspective from their own experiences with qualitative methods.

Analysis: We used template analysis,[25] a thematic analytic technique aligned with the critical realist epistemology that facilitates inductive coding of themes. After reading the first few transcripts, JLS developed an initial coding template, which the authors reviewed in relationship to the data and modified to develop an agreed upon set of codes. Three authors (JLS, CKB, JLB) applied this to two transcripts, and again discussed differences and modified code book definitions. The authors met numerous times throughout coding and made further minor modifications as needed. The authors often applied codes similarly; we used divergent coding as foci for discussion in refining the code book. All coding authors agreed on the final coding template (Supplemental Digital Appendix 2) and presentation of results (Chart 1). Two authors independently coded each transcript; JLS coded all transcripts; JLB and CKB each coded half of the transcripts using Dedoose (SocioCultural Research Consultants, LLC, Manhattan Beach, CA).

Together, the coding authors used template analysis to analyze coded excerpts and developed cohesive themes. We characterized challenges as related to the trainee, task, teacher, and setting, because previous research in procedural learning identifies these aspects as relevant.[18] We used the CLT framework to organize teaching strategies as those expected to match intrinsic load to the trainee's level, minimize extraneous load, or optimize germane load. Excerpts were selected to highlight themes and were labeled with subject number and abbreviations indicating specialty (GI, adult gastroenterologist; P-GI, pediatric gastroenterologist; AE, advanced endoscopist; SU, colorectal surgeon; FM, family medicine physician).

Ethical considerations: The University of California San Francisco Committee on Human Research and the Netherlands Association for Medical Education Ethical Review Board both reviewed and approved the study.

RESULTS

We interviewed 22 participants, including 8 adult gastroenterologists, 2 pediatric gastroenterologists, 6 advanced endoscopists, 3 surgeons and 3 family medicine physicians (Table 1). 7 (31.8%) participants were female. Participants reported an average 12.6 years (standard deviation 6.4 years) teaching endoscopy. Their learners included gastroenterology fellows (17, 77.3%), surgery residents (7, 31.8%), family medicine residents (3, 13.6%), nurse practitioners (3, 13.6%), practicing physicians (2, 9.1%), and colorectal surgery fellows (1, 4.5%). Altogether, 13.1 hours of interviews yielded 277 pages of transcripts.

Table 1. Characteristics of 22 experienced endoscopy teachers who participated in interviews regarding learner challenges and teaching strategies during gastrointestinal endoscopy training.

Characteristic	Descriptive statistic
Location, No. (%)	
United States	16 (72.7)
Netherlands	5 (22.7)
Canada	1 (4.5)
Clinical specialty, No. (%)	
Adult gastroenterologist	8 (36.4)
Advanced endoscopist[a]	6 (27.3)
Pediatric gastroenterologist	2 (9.1)
Colorectal surgeon	3 (13.6)
Family physician	3 (13.6)
Female gender, No. (%)	7 (31.8)
Years teaching endoscopy, mean (SD); range	12.6 (6.4); 4-30
Learners taught, No. (%)	
Gastroenterology fellows	17 (77.3)
Surgical residents	7 (31.8)
Family medicine residents	3 (13.6)
Nurse practitioners	3 (13.6)
Practicing physicians	2 (9.1)
Colorectal surgery fellows	1 (4.5)
Half-days of endoscopy performed per week, mean (SD); range	3.5 (2.3); 0.5-10
Proportion of endoscopy time spent teaching, mean (SD); range	0.60 (0.30), 0.08-1.0
Number of trainees in program per year, mean (SD); range	4.6 (1.9); 1-6

[a]Gastroenterologist who performs specialized procedures that are higher in risk and require additional training

Participants stories fit into three primary themes: (1) “diagnosing” trainees’ competence, (2) learning challenges, and (3) teaching strategies to promote learning. Chart 1 provides a summary of findings.

Chart 1. Summary of findings of analyzing transcripts from interviews with 22 experienced endoscopy teachers regarding learner challenges and teaching strategies during gastrointestinal endoscopy training.

Diagnosing the trainee	Learning challenges	Teaching strategies
Assess trainee skill Check trainee understanding Identify struggling trainees	*Related to the trainee* Performance goal-orientation Low self-efficacy Lack of awareness *Related to the task* Psychomotor control Tactile feedback Mental model *Related to teachers* Relationship with teacher Inadequate teaching skills Variability in teaching *Related to the setting* Distractions (internal and external) System structure	*To match intrinsic load to learner level* Teach along developmental continuum Provide motor instruction Provide technical assistance including take over *To minimize extraneous load* Optimize the environment Systems solutions Provide emotional support Communicate expectations and roles *To optimize germane load* Teach schemas Tell learner to stop and focus Promote a mastery goal-orientation

Diagnosing the trainee

Participants identified three approaches to “diagnose trainees” to estimate competence, characterize learning challenges, and target teaching strategies. They *assessed trainee skill* to determine what portions of the procedure a trainee could perform independently. Pre-session, teachers asked trainees about experience, progress and challenges to date, and occasionally reviewed results of formative and summative assessments. During procedures, participants assessed skills by observing performance quietly without intervening. “I'm assessing them while they're going towards the cecum. And the main way I assess is whether or not we're making forward progress and how much they're actually seeing the lumen and advancing versus just trying to guess where things are.” (S17, SU) Participants *checked trainee understanding* by asking probing and clarifying questions for rationale and reasoning, rather than immediately telling them what to do. “[I ask them], ‘What is the issue? How would you solve this? How would you approach it?’" (S1, GI)

Participants indicated that assessing skill and checking understanding helped them determine what portions of the task trainees could be entrusted to perform. This was most pertinent when participants had not recently worked with the trainee. “Whenever I start with a learner, when I haven't been with them for a while … the first procedure or two I just observe… At some point, if they really get into trouble I will deconstruct it and do it myself, but I like to … see what their method is.” (S2, GI)

Participants *identified struggling trainees* by observing body language, utterances, and responses to instructions or patient discomfort. “I can tell when they're overwhelmed because they're not having conversations about other things. They are trying the same thing over and over … they tend to stiffen up on their grip and become silent. You see them start trying to move the knobs, but you can tell it's not purposeful movement.” (S9, P-GI) Participants tried to determine when trainee struggling became unproductive, at which point they would intervene. “The main thing I try to balance [is] to figure out when … they're learning from the struggle, versus when they've crossed over into the point where they're struggling so much, they're just not learning from what they're doing.” (S17, SU)

Challenges

We categorized challenges as relating to trainees, tasks, teachers or settings, as follows.

Trainee-related challenges

Three challenges related to characteristics of trainees. Trainee *goal-orientation* influenced their ability to learn. Many were described as singularly focused on quick independent completion of the procedure (i.e., performance oriented), paying less attention to performing the procedure well, or learning from the experience.

> They [fellows] really want to go from the beginning to the end, they want to do the whole thing themselves. And it's … a point of pride that the attending should not have to touch the scope. … The goal in the fellow’s mind is that they have done the entire procedure themselves. However long it took, and whatever issues they encountered, that they somehow got … out of it themselves. (S16, P-GI)

This goal-orientation was also manifest by lack of interest in deeper cognitive learning or psychomotor practice on simulators. “They don't … always put in the time on the simulators and they don't … wanna take the time to learn the science behind what we're doing.” (S22, FM)

Low *self-efficacy* was thought to interfere with trainees’ receptivity to learning, which participants detected by trainees’ responses to directions or their body language, for example: “…you can tell [when] someone's shoulders droop [that] they've just given up and they don't really want to keep doing it and the sense of huge relief when I say, ‘Hey, do you want me to help you do it?’" (S3, GI)

Furthermore, some trainees seemed *unaware* of teachers’ instructions, or patient comfort and safety. This occurred when task complexity overwhelmed them.

> …they are too focused on what they are doing with their hands and not noticing what's happening elsewhere. I once witnessed a … technically competent fellow doing a colonoscopy, and the patient had … [an oxygen] desaturation down to 60%... People were getting nervous … and the [trainee], he [did not] notice. (S4, AE)

Task-related challenges

Three challenges related to the task. Most prominent was *psychomotor control* of the equipment; this was most common among earlier trainees.

> Probably the toughest skill for them to acquire … is fine tip control. Specifically, when they're trying to target a lesion or do some kind of therapy … and the stomach's moving or the diaphragm's going up and down. Who's going to be a great … fellow and who's not? The ones who … are in the process of mastering tip control are the ones that set themselves apart. (S11, AE)

Trainees were reported to have difficulty interpreting and acting upon *tactile feedback* from the endoscopic equipment. "It was mostly a manipulation issue with managing the scope and trying to get the loops out and knowing how much force to put in that the patient could tolerate without running the risk of perforating." (S17, SU) Participants also said trainees had difficulty developing a *mental model* of the physical configuration of the equipment inside patients' bodies. "Loops are another huge challenge for trainees because they're not often taught what different loops look like and why you're trying different things to correct them… They don't understand what's actually going on in the patient." (S9, P-GI) This was considered particularly challenging due to visual and tactile limitations and was contrasted with surgical teaching.

> The main difference … is, with surgical learning, you see what you're doing, you see the effects of the instrumentation. So, when you move something it instantly moves. When you touch something, you're actually touching it. And with endoscopy, you have no clue what that thing is doing on the inside. And if you did, it'd probably give you some palpitations. So, the thing is, it's just discordance. …you're watching a monitor and it's hard to know exactly what's going on inside. (S17, SU)

Teacher-related challenges

Three challenges related to endoscopy teachers. Participants reported their *relationship* with the trainee, as the teacher in the endoscopy room, led to trainee anxiety, distraction, or stress. "He was encountering problems with looping. It was looping at every segment, so it wasn't progressing … I told him to complete the endoscopy, so he was also stressed of course [that] I was looking. That was agitating him." (S5, AE)

Some participants reported that other teachers' inadequate *teaching skills* negatively impacted learning. "I scoped with a third-year fellow … and the way she was gripping her scope and handling the esophagus... I was astonished. It was clear to me that somewhere along the way people stopped trying to correct what was … bad technique and was going to be difficult for her to do anything in the future.” (S13, AE)

Participants thought that *variability* amongst teachers’ endoscopic approaches was confusing for early trainees. “I see fellows get frustrated coming from site to site, and they'll say, ‘This attending told me to do this and the next attending will be like, That's horrible! Why would anyone tell you that? You have to do the exact opposite.’” (S3, GI) However, some reported variability as beneficial for more advanced trainees, because different teachers taught different approaches or strategies that advanced trainees could understand and add to their repertoire. “I think variability is good … for learners who've had a certain amount of experience. Variability teaches them techniques and offers … an array of things that they can pull from their tool kit.” (S1, GI)

Setting-related challenges

Two challenges related to the setting. *Distractions* within the endoscopy room included people, pagers, tangential conversations, and internal distractions. “The interns' pagers are going off all the time… If they're … floating in and out during the day… they're more distracted versus if they're assigned to be there all day and not floating in and out.” (S19, SU) The *system structure* introduced competing demands that participants perceived as negatively impacting trainees’ ability to engage and learn.

> Someone's trying to do a procedure and their pager is going off repeatedly. You can see the demeanor changing, the focus. They're still looking at the screen, but their eyes look worried, like, “Oh no, it's another consult.” I think the way our program is set up … they're balancing an inpatient service and running back and forth and also going to clinic and then trying to squeeze in procedures. (S3, GI)

Teaching strategies

Teaching strategies were characterized as those expected to match intrinsic load to the learner’s level, minimize extraneous load or optimize germane load.

Matching intrinsic load

Participants described three teaching strategies that could align intrinsic load with trainees’ experience to support completion of the procedure. They adjusted trainees’ participation along a *developmental continuum*; earlier trainees would perform less while receiving more verbal support, whereas more advanced trainees could do more with less direction.

> The more advanced they are, the more competent they are, I'll give them much greater autonomy. [For the most] experienced, I'll let them start from the beginning; let them go around some turns and just offer some suggestions. If they’re less experienced, I'll insert and let them withdraw versus, [if they have] hardly any experience, I'll insert and withdraw most of the way and then just let them do it at the end. (S19, SU)

Some participants said their degree of trust in the learner impacted how they taught or how much they allowed the learner to do.

> When [they are] more senior and I have a sense of their skillset and I can trust them to think it through, I say, "Okay you are holding the scope, what do you wanna do?" Then we may have a discussion if we really diverge significantly on our ideas of what should be done, but … as they get better and better, I will give less instruction. (S14, GI)

Participants commonly reported using directive *motor instruction* to help trainees understand what to do or how to do something, for example: "You've got to make the scope drive around it, as opposed to getting up to the turn, slightly hooking it, torqueing the scope sideways to push the fold out of your way." (S11, AE) When trainees struggled despite instruction, participants provided *technical assistance or took over*. Technical assistance involved help with part of the procedure. "My approach has been to start with having them use the scope [shaft], not the dials. … To get them used to [the dials] and with me pushing initially. … It makes them able to focus … on … hardwiring the scope itself, the dials, the knobs, and the buttons." (S20, FM) Reasons for complete takeover included patient safety and comfort, time pressures, and the need to physically feel the instrument to understand what was happening. Most participants described thoughtfully considering when to takeover. "Almost always I allow them to stumble to the point of falling. And then [when] we hit a stalemate, I'll rescue them, so to speak. But we'll try to do every maneuver possible to allow them to accomplish it." (S22, FM) To make "takeover" time educational, participants described narrating maneuvers and thought processes while completing the procedure, and summarizing afterwards.

Minimizing extraneous load

Participants described four strategies that could reduce extraneous load. To *optimize the environment*, participants reported curtailing unnecessary conversations, silencing pagers, and limiting people in the room. *Systems solutions* were to adjust schedules, minimize competing demands, and leverage benefits of simulation-based learning. To reduce internal distractions, participants provided *emotional support* by encouraging trainees, and by identifying with and normalizing their struggles

> When he got into the stomach safely I said, "All right, drop your shoulders. Take a deep breath. … We're going to walk through the steps, but you're in control. I'm right here next to you. Everything is really stable." I spend a lot more time reassuring them… I find that some of what you're doing is fellow anxiety relief. (S13, AE)

Participants also reported that discussing *expectations and roles* could reduce trainee stress (which can induce extraneous load) related to uncertainty about their expected performance. "I expect it to take them a long time when they start… I say to them, 'As you get more experience, that will change. Right now, your goal is not to get to the end. Your goal is to [do the procedure] in a safe manner… it's okay if you can't finish the case.'" (S12, GI)

Optimizing germane load

Participants articulated three strategies that could promote germane load. Most often, they helped trainees understand cognitive and psychomotor concepts by *teaching schemas*. Approaches included verbal descriptions, demonstrating maneuvers outside of the patient, drawing pictures, showing appropriate amount of physical force, promoting practice of individual skills, using simulators, and leveraging technology (e.g., Google images, YouTube, specialized teaching equipment). "… another simple task that they struggle with in the beginning is just intubating the esophagus. I'm like, 'You're in the right position. This is the right amount of pressure.' I actually push on their shoulder, like 'This is how hard you should be pushing the scope.'" (S2, GI)

Surgeons discussed the unique ability to show trainees what was happening internally when the abdomen was open during intraoperative endoscopy.

> When I do intraoperative scopes… you can actually watch [the colonoscope] go through the colon. … I show the residents you can literally read the numbers on the scope through the colon wall just to impress upon them how tenuous things can be and how lucky it is that we don't put holes in more people. (S17, SU)

Some participants reported that when trainees appeared cognitively overloaded (e.g., when they repeated unsuccessful maneuvers, failed to progress, or exhibited frustrated body language or verbal utterances), they would instruct the trainee to *stop and focus* on their instruction. "I say, 'Stop. Here's what you're doing wrong. … You're going to hurt this person, so here's how you do this properly.'" (S18, SU) They also described *promoting a mastery orientation* valuing learning over performance. "They're so goal oriented… I'm not that impressed by you getting to the cecum quickly if you didn't do a good job. I'm very impressed if you go through [a] really hard sigmoid and then later on it's like super, super hard and I have to do it, but you did a great job with that." (S2, GI)

DISCUSSION

We analyzed stories told by experienced teachers of gastrointestinal endoscopy, a class of medical procedures in which the amount and fidelity of visual and tactile information is constrained. Our goal was to characterize learning challenges these experienced educators witnessed among their trainees, and teaching strategies they used to promote learning. Participants reported purposeful observation of trainees to assess skill and comprehension. They described learning challenges related to learners, tasks, teachers and settings. They articulated teaching strategies that aligned well with the CLT framework, providing practical guidance for teachers of complex medical procedures. In this section, we discuss our findings in the context of other models of procedural learning, practical implications for procedural skills training, and systems-related issues.

Procedural learning frameworks

Based on our prior research, we planned to use CLT as our primary theoretical framework, and CLT fit well with our participants' stories. Without prompting, most stories related to trainees who were cognitively overloaded, usually related to inherent task demands (i.e., intrinsic load) but sometimes related to distractions or competing demands (i.e., extraneous load). Likewise, many stories were about trainees who had difficulty understanding complex topics (i.e., germane load). We found that each teaching strategy fit intuitively with one of the three primary 'goals' of CLT. Despite the complexity of workplace learning, little CLT research in workplaces has utilized qualitative methodology. Rather, the vast majority has been quantitative with limited exploration beyond overall cognitive load or mental effort.[8] This limits conceptualization and understanding of how CLT can inform complexities of HPE, particularly in workplace settings, and we appreciate a need for further qualitative and mixed methods research linking CLT with procedural skills training.

Although we found CLT to be highly applicable, other frameworks are relevant. Fitts and Posner described three stages of psychomotor skills learning[26] that trainees progress through during procedural skills training,[27] including gastrointestinal endoscopy.[20] During the cognitive phase, trainees learn what needs to be done; they practice individual tasks and begin to learn the overall procedure in a step-by-step fashion. During the integrative or associative phase, trainees develop greater fluidity and automaticity for individual tasks and progressively accomplish entire procedures. In the autonomous phase, trainees develop competence to autonomously complete full procedures with accuracy and consistency. The three task-related challenges we identified link most directly to Fitts' and Posner's stages; these are critical challenges that impose intrinsic load. However, since trainee-, teacher-, and setting-related challenges can impose extraneous load, they may negatively impact trainees' progression through Fitts' and Posner's stages.

Ericsson's deliberate practice for mastery learning[28] is also relevant to procedural learning.[29] Participants referenced three essential features of deliberate practice[30] (appropriate level of difficulty, providing immediate informative feedback, trainee monitoring of understanding); no participant cited properties of mastery learning.[30] Rather, when participants discussed mastery, they referred to a mastery goal-orientation (focus on learning and improvement) versus performance goal-orientation (focus on achievement and demonstrating competence), as described by achievement goal theory.[31] We suspect two factors contribute to limited presence of Ericsson's principles in our interviews. First, most stories were of early trainees who were focused on the basics and were far from attaining mastery. Second, Ericsson's work primarily focuses on how practice is organized and performed over the course of training, whereas we focused on individual teaching encounters.

The physical nature of procedural learning implies cognitive activities are embodied within physical environments, so theories of embodied cognition are relevant. These theories hold that, because of the interconnectedness of the mind, body and environment, bodily experiences influence cognitive processes.[3] In other words, the mind does not process cognitive information in a vacuum, but, rather, within the context of sensorimotor experiences that occur within physical learning environments.[32] We suspect the task-related learning challenges our participants described are challenges of embodied cognition, rather than purely cognitive challenges.

The concept of multi-channel information processing is also pertinent. Mayer's cognitive theory of multimedia learning states that learning concurrently through visual and auditory channels can be effective, yet can induce cognitive overload.[33] In procedural learning, the tactile channel imposes a third modality for processing information, further increasing complexity and risk of cognitive overload, and many participants emphasized tactile challenges and strategies. Future study of interactions amongst the sensory channels and cognitive load during procedural learning would be informative.

Finally, some challenges relate to sociocultural models of learning.[34] A trainee's desire to "look good" for their supervisor could promote a performance over mastery goal orientation, and poor performance could induce low self-efficacy. External and internal distractions alike relate to the social context, as do teaching strategies including emotional support and communicating expectations and roles. We suggest elements of the social context could influence both extraneous and germane load.

Practical implications

Despite complexity of gastrointestinal endoscopy and its public health impact,[5,35] study of how trainees learn to perform endoscopy remains limited.[36] Most research has focused on learning curves,[37] assessing performance and competence,[9,38,39] trainee polyp detection,[40] and impact of simulation.[12] By identifying specific challenges and practical strategies to promote learning, our study advances understanding of teaching in this procedural field.

However, our study findings are relevant to teaching procedures beyond endoscopy. Gastrointestinal endoscopy is a "middle-ground" procedure in terms of complexity, risk, and duration, which we consider an ideal exemplar setting for studying procedural skills learning.[18,19] Teachers in any procedural setting could consider the framework of intentionally observing to diagnose learner skill, identifying specific learning challenges, and applying targeted teaching strategies.

Interpreting tactile feedback and developing mental models are among the greatest practical challenges when learning sensory-constrained procedures; these are challenges of schema formation. For early learners, schema teaching may be enhanced through simulation which allows the teacher to control intrinsic and extraneous load.[41] Indeed, simulation-based learning is effective in procedural teaching and HPE in general.[12,42] Our data suggest teachers may promote complex schema formation by using multiple teaching modalities as described by participants. A multimodal approach probably promotes learning more effectively than verbal teaching alone.[43] For example, some participants referred to the "ScopeGuide." This product allows learners to compare their mental model of colonoscope configuration with a cartoon depiction of the actual configuration in real time and has been shown to improve performance (though learning was not assessed).[5] Similar approaches could promote schema formation and learning in other procedural settings.

The most 'extreme' teaching strategy, complete takeover, reflects conflicting needs of multiple stakeholders.[44] The trainee needs to learn and develop autonomy. The patient needs a safe and high-quality procedure. The healthcare system demands efficiency. The teacher may feel impatient or worried about remaining procedures to complete. When patient safety and comfort are at stake, teachers should take over, but balancing other demands may present more ambiguity; indeed, several participants described conflicting feelings about whether and when to take over. Limited research suggests trainees feel frustrated when their supervisor takes over.[45-47] Setting expectations beforehand might reduce conflict and promote learning during takeover time. However, the subject would benefit from future research.

Systems issues are broadly relevant across procedural training programs. Competing demands inherent in clinical schedules were perceived to induce extraneous load by distracting the learner externally (e.g., pager beeping) and internally (e.g., worry about consults needing to be seen). Some participants described programmatic changes to reduce distractions, including dedicated procedural time without competing demands, and altering rotation and call schedules. Conversely, others felt distractions were authentic to actual practice and that such experience could therefore benefit trainees. It may be beneficial to shield early trainees in the partial-task phase from such competing demands, as such stimuli would mostly contribute to extraneous load. When learners can reliably complete whole tasks, such factors could be integrated as intrinsic demands of authentic workplace practice. Considering systems challenges more broadly, flexible and innovative approaches to trainee scheduling in procedural training programs may benefit trainees' learning. Additionally, teaching learners to metacognitively recognize and deprioritize distractions may be beneficial.[48]

Variability in teaching and lack of adequate teacher training are also challenges with systematic ramifications. While procedural learners desire less variability in teaching,[47] several participants perceived variability as beneficial for advanced learners. For earlier learners, consistency in teaching approach might reduce extraneous load. For example, Coderre et al[20] suggest using standard terminology for motor instruction, and dividing the procedure into standard portions. We agree with Coderre[20] that teaching procedural instructors (and we would add learners) about CLT is likely beneficial. We additionally recommend recognizing, articulating, and acting upon learner cognitive overload to address learning challenges.

As compared with cognitive learning, procedural learning often involves greater direct supervision. Of our 22 participants, 14 indicated trainees were directly supervised throughout every procedure, and the other 8 described ongoing supervision once trainees had established competence, albeit less intensive. Direct supervision may facilitate entrustment decisions with less external control or less intensive supervision.[49,50] Although interview questions did not ask about trust or entrustment, participant narratives suggested entrustment decisions were informed by assessing skill and checking understanding, and that level of trust impacted teaching along a developmental continuum. Although many factors should inform entrustment decisions,[51-53] many participants referred to direct observation of psychomotor skill and cognitive decision-making, whereas only two mentioned utilizing results of summative evaluations and competence committees. This likely owes to the focus of the interviews; had we asked about entrustment more directly, other factors might have been discussed. Our findings support the critical value of direct observation to inform entrustment decisions.[53]

Limitations

We studied a single group of procedures, yet our findings appear relevant to diverse procedural skills. We identified learning challenges by examining perspectives of teachers, rather than trainees. Because teachers cannot enter the minds of their trainees, their perceptions might differ from learners' perceptions. We chose the teacher's perspective because they have a deep understanding of how to perform and teach procedures (i.e., they have deep pedagogical knowledge[22]). Future research from the trainee perspective could provide a complementary view, and our findings could be useful in designing such research. The interviews did not enable us to directly link specific challenges with specific teaching strategies; while this was not our intent, future studies of when specific strategies are most effective would be helpful.

CONCLUSION

We identified procedural learning challenges and teaching strategies by analyzing stories that endoscopy teachers told about their experiences teaching learners to perform gastrointestinal endoscopy, and interpreted these findings through the lens of CLT. Further consideration and application of our findings may inform future research and direct teaching of procedural skills in the health professions.

REFERENCES

1. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci. 1988;12:257-285.
2. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. Med Teach. 2014;36(5):371-384.
3. van der Schaaf M, Bakker A, Ten Cate O. When I say ... embodied cognition. Med Educ. 2019;53(3):219-220.
4. Buchs NC, Pugin F, Volonte F, Morel P. Learning tools and simulation in robotic surgery: state of the art. World J Surg. 2013;37(12):2812-2819.
5. Kaltenbach T, Leung C, Wu K, Yan K, Friedland S, Soetikno R. Use of the colonoscope training model with the colonoscope 3D imaging probe improved trainee colonoscopy performance: a pilot study. Dig Dis Sci. 2011;56(5):1496-1502.
6. Spier BJ, Benson M, Pfau PR, Nelligan G, Lucey MR, Gaumnitz EA. Colonoscopy training in gastroenterology fellowships: determining competence. Gastrointest Endosc. 2010;71(2):319-324.
7. Canopy E, Evans M, Boehler M, Roberts N, Sanfey H, Mellinger J. Interdisciplinary cognitive task analysis: a strategy to develop a comprehensive endoscopic retrograde cholangiopancreatography protocol for use in fellowship training. Am J Surg. 2015;210(4):710-714.
8. Walker J, von Bergmann H. Lessons from a pilot project in cognitive task analysis: the potential role of intermediates in preclinical teaching in dental education. J Dent Educ. 2015;79(3):286-294.
9. ASGE Training Committee, Sedlack RE, Coyle WJ, et al. ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD. Gastrointest Endosc. 2014;79(1):1-7.
10. Vassiliou MC, Dunkin BJ, Marks JM, Fried GM. FLS and FES: comprehensive models of training and assessment. Surg Clin North Am. 2010;90(3):535-558.
11. Ritter EM, Cox TC, Trinca KD, Pearl JP. Simulated Colonoscopy Objective Performance Evaluation (SCOPE): a non-computer-based tool for assessment of endoscopic skills. Surg Endosc. 2013;27(11):4073-4080.
12. Singh S, Sedlack RE, Cook DA. Effects of Simulation-Based Training in Gastrointestinal Endoscopy: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2014;12(10):1611-1623.e1614.
13. Ritter EM, Lineberry M, Hashimoto DA, et al. Simulation-based mastery learning significantly reduces gender differences on the Fundamentals of Endoscopic

Surgery performance exam. Surg Endosc. 2018;32(12):5006-5011.
14. Ritter EM, Taylor ZA, Wolf KR, et al. Simulation-based mastery learning for endoscopy using the endoscopy training system: a strategy to improve endoscopic skills and prepare for the fundamentals of endoscopic surgery (FES) manual skills exam. Surg Endosc. 2018;32(1):413-420.
15. Thompson CC, Jirapinyo P, Kumar N, et al. Development and initial validation of an endoscopic part-task training box. Endoscopy. 2014;46(9):735-744.
16. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. 1956. Psychological Review. 1994;101(2):343-352.
17. Sewell JL, Maggio LA, ten Cate O, van Gog T, Young JQ, O'Sullivan PS. Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions. Med Teach. 2018; Oct 17:1-15 [Epub ahead of print].
18. Sewell JL, Boscardin CK, Young JQ, Ten Cate O, O'Sullivan PS. 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. 2017;92(11):1622-1631.
19. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. Med Educ. 2016;50(6):682-692.
20. Coderre S, Anderson J, Rostom A, McLaughlin K. Training the endoscopy trainer: from general principles to specific concepts. Can J Gastroenterol. 2010;24(12):700-704.
21. Maxwell JA. A Realist Approach for Qualitative Research. 1st ed: SAGE Publications, Inc.; 2012.
22. Shulman LS. Those Who Understand: Knowledge Growth in Teaching. Educational Researcher. 1986;15(2):4-14.
23. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. Qual Health Res. 2016;26(13):1753-1760.
24. Flanagan JC. The critical incident technique. Psychol Bull. 1954;51(4):327-358.
25. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psych. 2015;12:202-222.
26. Fitts PM, Posner MI. Human Performance. Oxford: Brooks & Cole; 1967.
27. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. N Engl J Med. 2006;355(25):2664-2669.
28. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and

related domains. Acad Med. 2004;79(10 Suppl):S70-81.
29. Griswold-Theodorson S, Ponnuru S, Dong C, Szyld D, Reed T, McGaghie WC. Beyond the simulation laboratory: a realist synthesis review of clinical outcomes of simulation-based mastery learning. Acad Med. 2015;90(11):1553-1560.
30. McGaghie WC, Kristopaitis T. Deliberate practice and mastery learning: origins of expert medical performance. In: Cleland J, Durning SJ, eds. Researching Medical Education. Vol I. West Sussex, UK: John Wiley & Sons, Ltd; 2015:219-230.
31. Dweck CS, Leggett EL. A social-cognitive approach to motivation and personality. Psychological Review. 1988;95(2):256-273.
32. Galetzka C. The Story So Far: How Embodied Cognition Advances Our Understanding of Meaning-Making. Front Psychol. 2017;8:1315.
33. Mayer RE. Multimedia learning. 2nd ed. ed: Cambridge University Press, New York, NY; 2009.
34. Kahlke R, Bates J, Nimmon L. When I say … sociocultural learning theory. Medical Education. 2019;53(2):117-118.
35. Gellad ZF, Provenzale D. Colorectal cancer: national and international perspective on the burden of disease and public health impact. Gastroenterology. 2010;138(6):2177-2190.
36. Thuraisingam AI, MacDonald J, Shaw IS. Insights into endoscopy training: a qualitative study of learning experience. Med Teach. 2006;28(5):453-459.
37. Siau K, Kuwai T, Ishaq S. Analysis of learning curves in gastroscopy training: the need for composite measures for defining competence. Gut. 2018;67(6):1198.
38. Kumar NL, Kugener G, Perencevich ML, Saltzman JR. The SAFE-T assessment tool: derivation and validation of a web-based application for point-of-care evaluation of gastroenterology fellow performance in colonoscopy. Gastrointest Endosc. 2018;87(1):262-269.
39. Sedlack RE, Coyle WJ. Assessment of competency in endoscopy: establishing and validating generalizable competency benchmarks for colonoscopy. Gastrointest Endosc. 2016;83(3):516-523.
40. Rustagi T. The influence of fellow involvement and level of fellowship training on adenoma detection rates. Gastrointest Endosc. 2010;72(5):1114-1115.
41. Fraser KL, Ayres P, Sweller J. Cognitive Load Theory for the Design of Medical Simulations. Simul Healthc. 2015;10(5):295-295-307.
42. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced

simulation: a systematic review and meta-analysis. Acad Med. 2013;88(8):1178-1186.
43. Shams L, Seitz AR. Benefits of multisensory learning. Trends in Cognitive Sciences. 2008;12(11):411-417.
44. Ten Cate O. Supervision and Entrustment in Clinical Training: Protecting Patients, Protecting Trainees. AHRQ Web M&M. 2018. Available online at https://psnet.ahrq.gov/webmm/case/461/Supervision-and-Entrustment-in-Clinical-training-Protecting-Patients-Protecting-Trainees. Accessed 12 April 2019.
45. Rose JS, Waibel BH, Schenarts PJ. Disparity between resident and faculty surgeons' perceptions of preoperative preparation, intraoperative teaching, and postoperative feedback. J Surg Educ. 2011;68(6):459-464.
46. Ko CY, Escarce JJ, Baker L, Sharp J, Guarino C. Predictors of surgery resident satisfaction with teaching by attendings: a national survey. Ann Surg. 2005;241(2):373-380.
47. Cooper MA, Tinmouth JM, Yong E, et al. Surgical residents' perceptions on learning gastrointestinal endoscopy: more hours and more standardization wanted. J Surg Educ. 2014;71(6):899-905.
48. Epstein RM, Siegel DJ, Silberman J. Self-monitoring in clinical practice: a challenge for medical educators. J Contin Educ Health Prof. 2008;28(1):5-13.
49. Santen SA, Wolff MS, Saxon K, Juneja N, Bassin B. Factors Affecting Entrustment and Autonomy in Emergency Medicine: "How much rope do I give them?". West J Emerg Med. 2019;20(1):58-63.
50. Salim SY, Govaerts M, White J. The Construction of Surgical Trust: How Surgeons Judge Residents' Readiness for Operative Independence. Ann Surg. 2018; Nov 29 [Epub ahead of print].
51. Duijn C, Welink LS, Bok HGJ, Ten Cate OTJ. When to trust our learners? Clinical teachers' perceptions of decision variables in the entrustment process. Perspect Med Educ. 2018;7(3):192-199.
52. Peters H, Holzhausen Y, Boscardin C, Ten Cate O, Chen HC. Twelve tips for the implementation of EPAs for assessment and entrustment decisions. Med Teach. 2017;39(8):802-807.
53. ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Acad Med. 2007;82(6):542-547.

SUPPLEMENTAL MATERIAL FOR CHAPTER 6.

Supplemental Digital Appendix 1. Interview Guide.

1. **[Specific example #1]**
   I would like to start by asking you about specific examples of endoscopy teaching. First, please think about an endoscopic procedure that you thought was particularly challenging for the learner. These challenges could relate to any portion of the procedure, including pre-, intra-, or post-procedure. Can you tell me about that encounter?
   [If needed can prompt with the following:]
   - Tell me about the learner (year, skill level)
   - Tell me about the procedure (what was happening, was it complex or routine)
   - In what ways, specifically, was the learner struggling?
   - Did anything appear to distract the learner?
   - How did you help the learner manage the situation?
   - Did you have any concerns about patient safety? Procedural efficiency?
   - Did you debrief or provide the learner with feedback after the procedure?

2. **[Specific example #2]**
   Thanks for that example. Can you think of another, different, situation in which a learner was struggling with a different type of endoscopic procedure? Can you tell me about that experience? How did the challenges they experienced differ from the first example? How did your teaching approach differ?

3. **[General challenges & approaches]**
   Thank you for describing those specific examples. Next, I am interested in discussing more generally the challenges that endoscopy learners commonly experience. Of course, I am also interested in any additional specific examples that you may have.
   [Depending on forgoing conversation and time, select among the following questions:]
   - What do you see as the most significant challenges that learners experience while learning to perform endoscopy (pre-, intra-, or post-procedure), and what techniques or approaches do you use to address these challenges?
     - Part task approach?
     - Setting expectations?
   - How, if at all, do you try to alter endoscopy teaching based on the fellow's prior experience or competence level?
   - How, if at all, do you try to minimize distractions in the environment, or how do you keep fellows focused on the task at hand?
   - How, if at all does fellow mindset affect their endoscopy learning?

- How, if at all, do you promote learning during endoscopy? By learning I mean automation of tasks, formation of learning schema, and consolidate knowledge.
- When you teach endoscopy do you tend to continuously narrate or do you tend more to sit back and observe?
- How do the challenges and/or your approaches to them change with advancing fellows skill and expertise?
- What procedures are most or least challenging for learners?
- Do fellows ever experience challenges pre- or post-procedurally (e.g., pre-procedure workup, report writing, follow-through)?
- How do you approach teaching about complex motor tasks, such as loops?
- How do you decide when to "take over" an endoscopic procedure? How do you decide whether or not to give the scope back to the learner? How do you keep learning active when you've taken over the procedure?
    - How does your level of comfort affect when or how you take over the procedure?
- How can you tell when a learner is distracted during endoscopy training? How do you help that learner?
- How can you tell when a learner is overwhelmed during endoscopy training? How do you help that learner?
- How does performance versus mastery mindset affect endoscopy learning?

4. **[Curricular & systems contributions]**
   Let's switch gears a little bit now to curricular and systems issues.
   In your training program, how do fellows learn to perform colonoscopy?
   Are there curricular or systemic factors that affect your fellows' endoscopy training in a positive or negative manner?
   [If needed can prompt with the following:]
    - Is there a consistent approach to endoscopy training across sites and/or teachers?
    - Does your program use simulation, and if so, how?
    - Is there a specific curriculum, and if so, can you describe it for me?
    - Do clinical or on-call schedules or competing demands interfere with endoscopy training?
    - Do fellows have access to adequate numbers of endoscopic procedures?
    - Does the endoscopy schedule facilitate enough time per case for fellows to learn?

Supplemental Digital Appendix 2. Definitions Of Codes.

Codes[a]	Definitions[b]
CHALLENGES	
Emotion/relationship	Learner emotion is source of distraction; includes emotions independent of, or related to, relationship/interaction between learner and teacher
Extrinsic challenges	Elements extrinsic to the learner, teacher, and patient are a challenge. Includes environmental elements like music, people, and conversations OR competing demands like consult service, pager, time pressure
Innate ability	Learner lacks innate skills required to learn endoscopy
Language challenge	Challenge understanding language or terminology used by teacher
Learner motivation/mindset	Performance over mastery orientation, cecum obsession, speed prioritized over quality, poor effort, lack of interest, lack of interest
Learner self-efficacy	Fear of falling back, lack of confidence, low self-efficacy
Mental model/loops	Learner doesn't CONCEPTUALLY understand/grasp what is happening with the equipment inside of the patient, including loops
Nonresponsive	Learner does not respond to feedback or instruction
Overall task	Complexity of overall task (as opposed to individual aspects) is overwhelming for learner
Psychomotor control	Any challenge related to physically controlling scope or tools including: - Controls (e.g., dials/knobs, suction, irrigation) - Maneuvering scope through body (e.g., intubating esophagus, positioning scope, MECHANICAL REMOVAL of loops) - Accomplishing tasks (e.g., biopsy, polypectomy, control of bleeding)
Systems problems	Systems or curricular issues interfere with learning, including: teacher inconsistency, inadequate practice and/or feedback, inadequate tools/supplies, inadequate curriculum
Tactile	Learner doesn't understand how scope should feel
Unawareness	Learner unaware of errors, struggles, own level of competence
STRATEGIES	
Check learner schemas	Teacher probes learner's understanding of complex concepts such as loops. Examples include asking learner to describe WHAT and WHY; asking the learner to propose and explain what is happening and what they want to do about it; asking learner to describe problem
Developmental continuum	Teacher proactively plans teaching based fellow's level in training and/or competence (e.g., part-task); infers that teacher permits more advanced learners to do more of procedure, and/or alters teaching based on experience/competence. Infers planned pedagogical approach, NOT real-time teaching adjustment

Diagnose struggler	Teacher observes learner IN REAL TIME to determine when they are struggling, including body language, utterances, posture, non-purposeful movements, making same maneuvers repeatedly
Emotional support	Teacher attends to learner emotions, such as: normalizes struggles, tells stories from own training, encourages, cheerleads
Expectation and roles	Teacher works with learner to set expectations for their roles during the procedure
Feedback	Includes any discussion of feedback
Mastery	Teacher encourages or promotes mastery mindset (over performance mindset), including demanding good technique
Motor instruction	Teacher teaches how to do specific motor tasks; examples: watch and imitate, narrate what teacher is doing, tell learner exactly what to do, demonstrate and then let learner try
Optimize environment	Teacher changes aspects of learning environment to improve teaching, such as leaving pagers outside, turning off music, minimizing conversation
Repetition	Teachers specifically gives learners same instructions over time, until they "get it"
Standard language	Teacher uses specific standardized language to make sure learner understands
Stop	Teacher tells learner to stop all movement so learner can listen to instruction
Systems solutions	Teacher or program makes changes to system or curriculum to improve learning. Examples include: learner/teacher continuity, scheduling dedicated endoscopy time free of competing demands, sequenced/scaffolded curriculum, faculty development, ALSO includes use of simulation
Takeover/technical assistance	Includes any discussion of takeover or when teacher physically assists with part of the procedure but does not completely take over
Teach schemas	Teacher helps learner understand abstract concepts such as how scope should feel and what is happening inside patient. Examples include: laying hands on scope together, hand motions, show scope on outside, push on learner body to show amount of pressure, draw pictures, use Google images, Scope Guide, analogies

[a]These are the codes that the coding authors used during actual coding. Coded excerpts were subjected to thematic analysis to develop the salient set of themes described in the Results section. Most codes can be found within the thematic analysis, but codes that were rarely applied (e.g., repetition, standard language) were not included.
[b]These operational definitions used by coding authors have been edited slightly for clarity.

Chapter 6

Trainee perception of cognitive load during observed faculty teaching of procedural skills

Sewell, JL

Young, JQ

Boscardin, CK

ten Cate, O

O'Sullivan, PS

Accepted for publication as:

Trainee perception of cognitive load during observed faculty teaching of procedural skills

Medical Education

ABSTRACT

Objectives: Although teachers impact learners' cognitive load, how specific teaching activities affect intrinsic, germane and extraneous load during procedural skills training is unknown. We sought to characterize teaching activities used in the exemplar procedural setting of colonoscopy, how they were enacted, and how learners perceived them as affecting intrinsic, germane and extraneous cognitive load.

Methods: We observed 10 colonoscopies performed by 8 different gastroenterology fellows and supervised by 10 different attending physicians at two hospitals, recording teaching activities observed, when they were used, and how they were enacted. After the colonoscopy, each fellow completed the Cognitive Load Inventory for Colonoscopy to quantify intrinsic, germane and extraneous load. We then interviewed each fellow to determine how they perceived teaching as affecting cognitive load. Qualitative data were subjected to content analysis. We correlated instances of germane load-promoting activities with measured germane load.

Results: We observed 515 instances of teaching activities. Intensity of teaching varied substantially, ranging from 0.7 to 3.3 activities per minute, as did pattern of teaching activities used by different attendings. Little teaching occurred immediately before or after procedures. Fellows usually perceived teaching as affecting cognitive load in ways that promoted learning, particularly reducing intrinsic load and increasing germane load. Fellows strongly perceived that providing autonomy promoted germane load. Conversely, fellows perceived that excessive teaching increased extraneous load. Instances of germane load-promoting teaching activities correlated moderately with measured germane load.

Conclusions: Teaching during the exemplar procedural setting of colonoscopy affected learners' cognitive load in mostly beneficial ways, yet even "good" teaching activities had detrimental effects when used excessively. Procedural teachers should consider learner experience, task complexity, and environmental factors to modulate modality, content and intensity of teaching to promote balanced cognitive load and learning. Teaching more reservedly during the procedure and taking advantage of pre- and post-procedure opportunities may help.

INTRODUCTION

Teaching and learning procedural skills in the health professions is challenging[1,2], since teachers and learners alike must balance demands of multiple stakeholders including patients, teachers, learners, other providers and staff, and healthcare systems. Procedural skills often involve substantial cognitive and psychomotor complexity, and potential for external and internal distractions, so risk for working memory overload is high, making cognitive load theory (CLT) relevant[3].

CLT posits working memory as the bottleneck for learning, because working memory can manage only few pieces of information at any moment (as opposed to sensory and long-term memory which are considered infinite)[4]. Three types of cognitive load impose on working memory. Cognitive load is categorized as *intrinsic load* when learners complete essential components of a learning task, as *extraneous load* when learners attend to non-essential task elements (including external distractions imposed by the environment and internal distractions related to their thoughts or emotional state), and as *germane load* when learners link information to create or refine learning schemas, which leads to automation and learning. Extraneous load should be minimized and intrinsic load matched to the learner's experience level; in the motivated learner, this provides working memory space for cognitive activities promoting germane load and learning[4]. While intrinsic and extraneous load are considered incontrovertible types of cognitive load, debate exists as to whether germane load is a unique form of cognitive load, or whether it is a component of intrinsic load[5-7]. Since germane load is the cognitive load type most directly linked to learning, educators should find it of great interest. However, we know little about factors or variables that impact germane load in health professions education (HPE) workplace settings, and it remains less-studied than intrinsic or extraneous load[8].

Empirical research confirms relevance of CLT to procedural training in HPE settings[8], yet how specific teaching activities may affect learners' intrinsic, germane and extraneous cognitive load remains uncertain. Procedural teachers undoubtedly intend their teaching to promote learning, yet some teaching approaches might have unintended consequences that induce extraneous load or cognitive overload, which interfere with learning. For example, a cardiac surgeon might intend her detailed description of valvular pathophysiology to help the surgical intern understand the rationale for the surgical approach, but if her teaching is delivered using vocabulary unfamiliar to the intern, or in the midst of a complex operation, the teacher's explanation could contribute to working memory overload and hinder the intern's learning and performance. Better understanding of how specific teaching approaches in procedural settings affect learners' cognitive load could help guide procedural teachers. We sought to address this gap by studying a specific genre of medical procedures: gastrointestinal (GI) endoscopy.

GI endoscopy encompasses a set of moderately complex and invasive procedures that provide significant benefits for populations such as lifesaving management of hemorrhage and diagnosis and prevention of colorectal cancer[9,10]. Learning to perform GI endoscopy is complex and challenging, in part because visual and tactile feedback are both constrained[10]. Trainees must complete several hundred of each type of endoscopic procedure, and master multiple ancillary maneuvers such as biopsy, polypectomy, and control of hemorrhage, before competence can be attained[11]. We and others have demonstrated relevance of CLT to GI endoscopy training[12-14], and we consider GI endoscopy as an exemplar setting in which to study procedural teaching and learning[12,13]. Better understanding of how teaching affects cognitive load could lead to improved procedural skills teaching, a better trained junior workforce, and improved outcomes for patients. Lessons learned in this exemplar setting should be generalizable to diverse HPE procedural learning settings.

Previously, we characterized teaching activities that experienced endoscopy teachers reported using when teaching GI endoscopy[15]. Participants in that study, who were purposively sampled to maximize experience and insight into educational processes, reported using 11 specific teaching activities. However, as participants were highly selected for their educational expertise and insight, we do not know whether ‘typical’ endoscopy teachers use these activities, how the activities are enacted, or how they affect learners’ cognitive load. In the current study, we seek to better understand these gaps by addressing three research questions:

1) Which teaching activities do endoscopy teachers actually use when teaching gastroenterology fellows to perform colonoscopy, how often do they use them, and how are they enacted?
2) How do fellows perceive those enactments as affecting intrinsic, germane and extraneous cognitive load, and are perceived effects consistent with those we predict based on CLT?
3) Do activities perceived to promote germane load correlate with measured germane load?

METHODS

Design, participants and procedures: We performed a concurrent transformative mixed methods[16] study (concurrent collection of quantitative and qualitative data that simultaneously inform the study aims). The lead author observed gastroenterology fellows performing colonoscopy under the supervision of attending gastroenterologists at a tertiary university hospital and a large urban safety net hospital. We utilized one, rather than two, observers for several reasons. First, colonoscopy is a sensitive and vulnerable procedure that exposes patients’ buttocks, anal and perineal regions, and we believed that two observers would pose excessive embarrassment for patients (for this reason we also did not video or audio record). Second, the small endoscopy rooms could not accommodate two observers without distracting attendings and fellows performing the procedure. Third, consistent with qualitative research a well-trained and knowledgeable observer is well-positioned to do the observation.

We purposively sampled attending physicians to include diverse years in practice, and both men and women. We purposively sampled fellows primarily within the first year of training; they have less experience than more senior fellows and therefore will need more teaching, thereby enriching the number and diversity of teaching activity enactments by attendings. However, endoscopy trainees in their first couple of months are able to do very little endoscopically, which could restrict teaching strategies. We therefore purposefully waited to recruit until a few months into the fellows' first year, at which point they would have participated in several dozen procedures and have greater ability to participate meaningfully. We recruited a smaller number of fellows in the second year of training, with the assumption that teaching of these more experienced learners might differ. We specifically did not recruit fellows in the third year of training; since they have developed so much automaticity, teaching required by the attending might be limited. We observed each attending only once, and each fellow no more than twice.

Observation commenced when the fellow, attending, and patient were all present in the room, and continued until the attending and fellow finished talking after completing the colonoscopy. During each observed colonoscopy, the lead author used event sampling to identify instances of 11 specific teaching activities derived from prior empirical research[15,17] (Table 1). We categorized these strategies *a priori* as those primarily predicted to reduce intrinsic load, increase germane load, or reduce extraneous load; we made secondary predictions for some of them. We focused on reducing, rather than optimizing intrinsic load; because colonoscopy is so complex, learners at the level of our participants usually require teaching and attending assistance to complete the procedure. The lead author recorded the portion of the procedure during which teaching activities were used (insertion, withdrawal or before/after procedure). The author transcribed words said (verbatim or paraphrases) and actions performed using an Excel form (Microsoft Corp., Redmond, WA) developed for this purpose. The author also recorded when attendings used teaching activities other than those in Table 1. The author practiced using this data collection method several times before actual observations commenced.

Table 1. Teaching activities observed for during colonoscopy.

Teaching activities	Definitions	Additional predicted cognitive load effects
Primarily predicted to decrease intrinsic load		
Motor instruction	Attending tells fellow specifically what to do with colonoscope or instruments	Increase germane load
Technical assistance	Attending assists with ancillary instruments (e.g., biopsy forceps, polypectomy snare), but fellow maintains primary control of colonoscope	None
Take over	Attending assumes primary control of colonoscope while fellow observes	Increase germane load
Primarily predicted to increase germane load		
Check learner understanding	Attending asks 'how or why' questions to check fellow's understanding of procedure	None
Feedback	Attending provides evaluative comments on fellow's performance, either specific or general	Decrease intrinsic load
Promote mastery mindset	Attending tells fellow to focus on doing or learning good technique (as opposed to just trying to complete the colonoscopy quickly)	None
Stop and focus	Attending tells fellow to stop motor activity and to focus on their instruction	Decrease extraneous load
Teach schemas	Attending explains what, how, or why, either related to the specific procedure or regarding colonoscopy in general	Decrease intrinsic load
Primarily predicted to decrease extraneous load		
Emotional support	Attending makes statements that could reduce fear, anxiety, worry, or negative emotions	Increase germane load
Set expectations	Attending describes roles of themselves and/or the fellow, or tells the fellow what they can expect during the procedure	Increase germane load
Modify environment	Attending modifies aspects of the environment (e.g., gurney, patient position, monitor, fellow stance), or instructs other persons to behave differently with intent to help fellow	Decrease intrinsic load

After the colonoscopy was complete, the fellow rated her/his degree of overall mental effort (using Paas' mental effort rating item[18], a 1-9 scale) and cognitive load subtypes (using the Cognitive Load Inventory for Colonoscopy [CLIC][13], a 0-10 scale). On both scales, a higher score indicates a higher degree of mental effort or cognitive load.

The lead author then performed a semi-structured interview with each fellow in a separate room. The colonoscopy experience was used as critical incident technique[19] to address the fellow's perceptions of the teaching that occurred and how teaching activities were perceived to affect their cognitive load. Interviews were audio recorded and transcribed. Primary interview questions were as follows.

1) How did the colonoscopy go for you? How easy or difficult was it? What went well or did not go well?
2) What teaching activities did your attending use to help you complete and learn from the colonoscopy? *[Prompt as needed with any observed teaching activities that fellow does not mention. If fellow recalls activity after prompting, then discuss. If fellow does not recall after prompting, do not discuss.]*
3) *[For each teaching activity observed]*: Did that activity either help or hinder your ability to complete the task, minimize distractions, or learn from the task?

With regard to question 2, when fellows described activities not immediately matching those in Table 1, the interviewer asked clarifying questions to ascertain whether it might represent one of the predetermined activities. If not, it was categorized as 'other' and discussed among authors to determine whether a new activity type was warranted. With regard to question 3, 'completing the task' referred to intrinsic load, 'minimize distractions' referred to extraneous load, and 'learn from the task' referred to germane load.

Quantitative data: Items from the CLIC assessing each type of cognitive load were averaged together to produce a single estimate of intrinsic, germane and extraneous load, as previously described[13]. These and other quantitative data were assembled in Excel.

Qualitative data: During the colonoscopy observation portion of the study, the observer categorized teaching activities as they occurred. If the type of teaching activity was not immediately apparent, it was recorded as 'other' and revisited later for categorization. As interviews were transcribed, two authors read each transcript and applied content analysis to code how fellows perceived each teaching activity as affecting intrinsic, germane and extraneous load, using a Qualtrics® form developed for this purpose (Supplemental Material). All transcripts were doubly coded; the lead author coded all, while three other authors (JQY, CKB, PSO'S) divided second coder duties. The lead author met with each other coding author to reconcile differences in coding by consensus. If consensus could not be reached, a third author helped reconcile. Due to limitations in numbers of faculty and fellows, we planned *a priori* to start with a sample of 10 observations and to assess at that time for information sufficiency[20] of teaching activity enactment, and to perform additional observations if novel enactments were still being observed.

Analysis: For research question 1, we used content analysis to categorize teaching activities used during colonoscopy observation, and to characterize enactment of those activities. We selected excerpts to highlight findings. To protect participant privacy, excerpts were labeled with randomly assigned letters and gender was redacted. We used descriptive statistics calculated in Excel to summarize how often specific teaching activities were used, in which portion of the colonoscopy there were used, and teaching patterns for each attending. We generated descriptive statistics for the cognitive load scores and other data.

For research question 2, we used content analysis to characterize perceived cognitive load impact and compared perceived impacts to our predictions.

Given the small sample size, we performed a single statistical test for research question 3. We identified teaching activities for which the most common perceived impact was to increase germane load. For each observation, we tallied the number of times these activities occurred. We then ran a Spearman rank-order correlation using Stata version 14 (StataCorp, College Station, TX) to examine the correlation between the number of times germane load-promoting activities were observed and the germane load score from the CLIC.

Reflexivity: The lead author is a gastroenterologist with more than 7 years' experience teaching endoscopic procedures; he has personally completed more than 4,000 colonoscopies and is an education researcher focused on CLT. Other authors have experience and expertise in HPE research and CLT.

Ethical considerations: The study was reviewed and approved by the University of California San Francisco Institutional Review Board.

RESULTS

We observed 10 colonoscopies involving 10 different attendings and 8 different fellows. Attendings reported a median 6 years of prior teaching experience (range 1 to >20 years). Fellows were primarily in the first year of training (8, 80%) and reported a median prior experience of 82 colonoscopies (range 50 to 442). Two fellows (25%) and 4 attendings (40%) were female. Median (range) procedural time was 33 minutes (22 to 50). Median (range) cognitive load (0-10 scale) was highest for germane load [4.5 (1.5-8.8)], followed by intrinsic load [3.0 (1.6-6.4)] and extraneous load [1.6 (0.0-3.8)]. Median (range) overall mental effort (1-9 scale) using Paas' scale[18] was 7 (4-9). Table 2 summarizes characteristics of participants and procedures.

Table 2. Characteristics of participants and procedures.

Characteristic	Descriptive statistic
Fellows (N=8)	
Year in training, No. (%)	
PGY-4	8 (80%)
PGY-5	2 (20%)
Female gender, No. (%)	2 (25%)
Cognitive load, Median (range)	
Intrinsic load*	3.0 (1.6-6.4)
Germane load*	4.5 (1.5-8.8)
Extraneous load*	1.6 (0.0-3.8)
Overall mental effort†	7 (4-9)
Attendings (N=10)	
Years teaching colonoscopy, No. (%)	
1-5	5 (50%)
6-10	3 (30%)
11-20	0 (0%)
>20	2 (20%)
Female gender, No. (%)	4 (40%)
Colonoscopies (N=10)	
Time to reach cecum in minutes, Median (range)	12 (5,26)
Total colonoscopy time in minutes, Median (range)	29 (22,50)
Number of teaching activities used per colonoscopy, Median (range)	39 (23,148)

*Measured using Cognitive Load Inventory for Colonoscopy[13], which could range from 0-10; higher scores indicate higher levels of intrinsic, germane or extraneous load. †Measured using the Paas Single Item Rating of Mental Effort[18], which could range from 1-9; higher score indicates higher level of overall mental effort.

We observed 515 instances of the 11 teaching activities during the colonoscopies. Supplemental Figure 1 depicts the frequency of observation for each teaching activity and the distribution of colonoscopic segments during which each was used. Most activities were distributed fairly equally between insertion and withdrawal phases of colonoscopy; few were used before or after colonoscopies. The most frequently used teaching activities were among those predicted to affect intrinsic and/or germane load: motor instruction, feedback, schema teaching, checking understanding, and takeover. Two strategies, "promote mastery mindset" and "stop and listen," were not observed during any colonoscopy.

Use of teaching activities varied widely among attendings. Median number of teaching activities per colonoscopy was 42 and ranged from 15-136. Median number of activities used per minute of procedure time was 1.0 per minute and also varied substantially from 0.7-3.3 per minute (Supplemental Figure 2). Numerically, the greatest variation was seen in frequency of motor instruction and feedback.

Teaching activities, enactments and perceived cognitive load impact: Each activity is discussed below, ordered by descending frequency of use (see Table 3). During the final few observations, we observed no new enactments of any teaching activity; we therefore concluded we had reached information sufficiency and did not perform additional observations. There were no apparent differences in perceived cognitive load impact of strategies between first- and second-year fellows.

Motor instruction was observed 215 times during 9 colonoscopies. The primary form of enactment was specific comments given orally that were intended as directives that the fellow should follow. Fellows perceived verbal motor instruction to decrease intrinsic load; because the attending assumed the cognitive task of deciding what to do, the fellow could focus on accomplishing motor tasks that the attending dictated. Two fellows perceived this increased germane load; in both cases the attending gave significantly less motor instruction (7 and 8 instances, respectively), and instructions were related to specific tasks with which the fellow was struggling. A less common enactment of motor instruction was physical gestures, such as pointing to direct attention or hand movements to show directionality or movement. Fellows reported not noticing the gestures and perceived no cognitive load impact. Two fellows perceived motor instruction to have adverse effects on cognitive load: decreased germane load and increased extraneous load. In both cases, the amount of verbal motor instruction was excessive and made the fellow feel overloaded. Fellow C said, "[The attending] gives instruction particularly quickly, so I'm not given my own time to problem-solve and figure things out on my own".

Feedback (observed 97 times during 9 colonoscopies) was enacted orally in two different ways: nonspecific and specific. Overall, the most common perceived effects on cognitive load were to decrease intrinsic load and increase germane load.

Nonspecific feedback, observed in 8 colonoscopies, included general comments like "Good job" or "Nice". Perceived cognitive load impact varied. Fellows took affirmative nonspecific feedback as a signal of doing a good job. Some fellows interpreted it meant they could reduce vigilance of colonic inspection (i.e., decreased intrinsic load). Others perceived it as confirming good endoscopic technique (i.e., increased germane load). Still others felt the feedback reduced anxiety, and therefore reduced extraneous load. Additionally, a few reported nonspecific feedback had no cognitive load impact. Finally, one fellow perceived nonspecific feedback to increase extraneous load; they thought the attending was subtly rushing them complete the colonoscopy more quickly.

Specific feedback, observed in 3 colonoscopies, included statements on what the fellow did well or did not do well, usually with directions on how to improve. Fellows perceived specific feedback helped them know what to do and thus decreased intrinsic load, and that it contributed to understanding what was happening and hence increased germane load. Feedback was perceived as more beneficial when it related to topics mentioned in pre-procedure discussions.

Schema teaching (observed 77 times during 9 colonoscopies) was usually enacted by orally describing general truisms or rules, or by explaining complex concepts. One attending performed a physical enactment by pushing on the fellow's arm to demonstrate the amount of force needed to complete a colonoscopic maneuver. The most common perceived impact was increased germane load. However, fellows reported schema teaching had no cognitive load impact when they already knew what the attending was trying to teach, or when the task was so complex that they could not pay attention to the attending's teaching.

Checking understanding (observed 50 times during 10 colonoscopies) was enacted by orally asking the fellow what s/he thought was happening, or what s/he wanted to do next, as opposed to directly telling the fellow this information. The primary effect was a perceived increase in germane load, by promoting critical thinking and active decision-making. When the fellow perceived they were "on the same page" as the attending this helped cement learning schemas. The attending's questions led one fellow to attempt solutions that they felt promoted completion of the colonoscopy (i.e., decreased intrinsic load). For two fellows, the attending asked questions that the fellow did not know how to answer, which they felt increased extraneous load. Fellow F said, "I think that [the attending asking clarifying questions] … added to the distraction… At that time I was still trying to find where the lumen was. So I think when [the attending] asked at that time it added extra things for me to consider. … I wasn't at a place where I could think of other things." Two other fellows did not notice the attending's questions and perceived no impact on cognitive load.

Activities related to *takeover* (observed 38 times during 7 colonoscopies) were enacted physically by the attending taking control over the colonoscopy when the fellow was having difficulty progressing or when there was time pressure. Most attendings described their motor approach, either during or after takeover. Fellows perceived takeover to decrease intrinsic load and increase germane load. Fellows reported that when the attending took over, this freed attentional capacity to listen to the attending's words and concentrate on what the attending was saying and doing in a way that was not possible when the fellow held the colonoscope. As Fellow B said, "I want to see how they do it. Then I get the instant gratification, like, 'Oh, that's how they do it.' Then I've got that in my pocket now."

Setting expectations (observed 12 times during 7 colonoscopies) was enacted orally, either by telling the fellow when the attending might take over the procedure (based on how much time or how many attempts the fellow would be permitted) or by discussing with the fellow what they wanted to focus on during the colonoscopy. The former enactment had less perceived impact; one fellow reported when they knew the attending would take over at a particular time, it reduced negative feelings when takeover occurred (i.e., decreased extraneous load), but other fellows perceived no impact. The latter enactment, however, was perceived to increase germane load; fellows reported this helped them know what to focus on during the colonoscopy, which promoted learning. Fellow B said, “It's nice because [the attending] holds you accountable. [The attending] asks you what you want to learn up front or what you're working on and then how you're going to measure it and then asks you to assess what your measurement of that skill was at the end … it makes you use each colonoscopy as a teaching experience as opposed to just trying to get to the cecum, move onto the next one.”

Technical assistance (observed 12 times during 7 colonoscopies) was enacted physically by passing and/or manipulating instruments to help the fellow (most practicing endoscopists pass and manipulate instruments autonomously). This activity primarily was perceived to decrease intrinsic load, as it reduced motor complexity for the fellow (i.e., part task approach). One fellow said that watching what the attending did with the instrument also promoted germane load, though another fellow felt that the attending’s assistance impaired their ability to learn how to use the instruments (i.e., decreased germane load).

Emotional support (observed 10 times during 5 colonoscopies) occurred when attendings identified with the fellow’s struggles. Fellows perceived this decreased extraneous load by diminishing shame or other negative emotions arising from not performing as well as desired. Fellow J said, “I think it evens the playing field a little bit and makes me realize that [the attending] remembers what it's like to be a learner”. One fellow reported this enabled them better able to hear and apply the attending’s teaching (i.e., increased germane load). However, another fellow was surprised when the attending asked about their emotional state, and felt this was a distraction (increased extraneous load).

Environment modification (observed 4 times during 4 colonoscopies) was enacted physically by adjusting the patient position, the bed or the monitor into a better position. Perceived effects on cognitive load were mixed; one fellow perceived decreased intrinsic load, another increased extraneous load, and two perceived no cognitive load impact.

Table 3. Enactments and cognitive load impacts.

Strategy, No. instances, No. colonoscopies[*]	Types of enactments with illustrative excerpts for verbal enactments[†]	Perceived effects on cognitive load (No. fellows)[‡]
Motor instruction **N=215, 9**	• Verbal o "Pull back a little and look up for the polyp." o "Now [tip] down and right, you gotta get over there, tip down." o "Turn a little using the little dial and tip up." • Physical: hand gestures	**IL↓ (8)** GL↑ (2) EL↑ (2) No impact (2) GL↓ (1)
Feedback **N=97, 9**	• Nonspecific statements: o "Nice" or "Good job" • Specific statements: o "You took advantage of going towards 12 o'clock since you get a lot better tip flexion there" o "It's good that you're not pushing too much, but use your dials here instead."	**IL↓ (4)** **GL↑ (3)** No impact (3) EL↓ (2) EL↑ (2)
Schema teaching **N=77, 10**	• "In the right colon I first examine the lower wall, then I go back and examine the upper wall." • "Whenever you can try to do it by pulling, not pushing." • "If you're up against the wall you're just hoping for the best. You gotta get off the wall."	**GL↑ (6)** No impact (4) EL ↑ (1)
Check understanding **N=50, 10**	• "Do you know where you need to go?" • "What are your recommendations for this patient? When should her next colonoscopy be? When can she restart aspirin?" • "How do you want to remove that polyp?"	**GL↑ (4)** IL↓ (2) EL↑(2) No impact (2)
Take over **N=38, 7**	• Physical: attendings took over most often when fellow was not progressing or when there were time demands • Verbal: attendings dialogued while they controlled colonoscope and summarized what they did after complete	**GL↑ (6)** **IL↓ (5)** EL↑(1)
Set expectations **N=12, 7**	• "As soon as I get around this turn you can take over again" • "I'll probably give you 10 minutes [to get to cecum]." • "Is there anything [during the colonoscopy] that you want me to specifically watch for?"	**GL↑ (3)** No impact (2) IL↓ (1) **EL↓ (1)** EL↑ (1)
Technical assistance **N=12, 7**	• Attendings passes the snare or biopsy forceps for the fellow • Attending pushes on patient's buttock to stop air from passing out	**IL↓ (4)** GL↑ (1) GL↓ (1)
Emotional support **N=10, 5**	• "Ok, that was not easy [for me either]!" • "I remember when I was a fellow that always happened."	**EL↓ (4)** GL↑ (1) EL↑ (1)
Modify environment **N=4, 4**	• Physical: attending adjust bed height or patient position • Verbal: asked about same things, e.g., "Do you want to bed a little lower?"	No impact (2) IL↓ (1) EL↑ (1)
Promote mastery mindset **N=0, 0**	Not observed	N/A
Stop and listen **N=0, 0**	Not observed	N/A

LEGEND FOR TABLE 3: EL, extraneous load; GL, germane load; IL, intrinsic load; N/A, not applicable; ↑, increased; ↓, decreased. [*]Presented in descending order of frequency of use. [†]Phrases in quotes are paraphrases of statements made by attendings during observations. Phrases not in quotes are actions observed. [‡]Bold font indicates cognitive load effect matched hypothesized effect (Table 1).

Other teaching activities: We identified two additional teaching activities: provision of cognitive information and permitting autonomy.

Cognitive information was first identified during the third observation and was observed during 7 colonoscopies. This activity was enacted by telling the fellow small pieces of information that were not connected to specific motor instructions or larger schemas. For example, Attending F said, "He's got a bit of a floppy colon," and Attending G said, "You're right outside the cecum". Cognitive information was primarily perceived to decrease intrinsic load by decreasing cognitive complexity, allowing the fellow to focus on motor portion of the task, and to increase germane load when the information related to what the fellow was focusing on or thinking about at the time. However, in one case where excessive cognitive information was given, the fellow felt overwhelmed and perceived increased extraneous load and decreased germane load.

Autonomy was first identified during the sixth observation and was observed during 4 colonoscopies. This activity was enacted physically by observing the fellow without commentary or verbal instructions; sometimes the attending would physically step away from the fellow. Fellows reported when they felt the attending was affording them autonomy, it promoted germane load. The more extended period of focused practice without attending intervention or interruption allowed them to appreciate consequences of their actions, to see what worked, and to feel a sense of ownership and agency. Fellow J said, "It [the attending affording autonomy] helped me learn … doing it myself … is the best way for me to learn about how the scope feels," and Fellow E said, "I really appreciate that [the attending] doesn't … get stressed about those details... It helps to learn from the procedure because you [can focus on] the bigger thing." Conversely, limited autonomy was perceived to negatively affect germane load. Fellow C said, "I think sometimes it [continuous narration without perceived autonomy] does interfere [with the ability to learn] because … I'm not given the opportunity to figure out what works best for me".

Germane load

Instances of the four teaching activities with a primary positive effect on germane load (schema teaching, takeover, checking understanding and setting expectations) had a positive correlation of moderate strength (Spearman's rho 0.44) with measured germane load. The scatterplot (Supplemental Figure 3) visually supports this correlation considering either first-year fellows alone or both first- and second-year fellows.

DISCUSSION

Although CLT is considered relevant to procedural skills training in HPE, little is known about how teaching affects the intrinsic, germane and extraneous cognitive load procedural learners experience[8]. In this study, we characterized teaching activities used during the exemplar setting of colonoscopy, how the activities were enacted, and what impact fellows perceived the teaching activities as having on cognitive load. Although frequency of teaching activity use varied substantially among the 10 colonoscopy teachers, their enactments of the activities were quite similar. The moderate correlation between activities perceived to promote germane load and measured germane load provides evidence supporting germane load as a construct. Teaching activities tended to have perceived effects that benefited learners and promoted goals of CLT, particularly to increase germane load and reduce intrinsic load. However, even “good” teaching activities were sometimes used in ways that interfered with learning, primarily by contributing to extraneous load. Our findings advance understanding of procedural skills training through the lens of CLT and suggest that concepts of CLT may be used to promote effective procedural teaching.

Very few studies have investigated the three types of cognitive load in procedural settings[8]. Our prior work quantitatively assessed characteristics of learners, teachers, tasks and settings that were associated with intrinsic, germane and extraneous load among colonoscopy learners[15]. Naismith et al qualitatively identified characteristics of a simulation that respirology and internal medicine residents perceived as contributing to intrinsic, germane and extraneous load while performing simulated pleural catheter drain placement, thoracentesis and paracentesis (they also compared quantitative cognitive load measures)[21]. While both of these studies produced practically and theoretically valuable findings, neither assessed how specific teaching affected cognitive load, and our current study complements and builds upon these.

Frequency and intensity of teaching

There was variability in the frequency with which particular teaching activities were used. Activities most commonly related to task completion (i.e., intrinsic load) and learning (i.e., germane load); activities related to extraneous load were used much less often. This was consistent with measured cognitive load; germane and intrinsic load were rated as greater than extraneous load. It only makes sense that some activities were used more than others, but we were surprised to observe no instances of either “promote mastery mindset” or “stop and listen,” since these were mentioned by participants in our prior study[15]. Participants in that study, who were purposively selected for deep education experience and insight, might be more likely than the ‘typical’ GI endoscopy teacher to use these more reflective teaching activities. The lack of observation does not imply these teaching strategies as being irrelevant, but does prevent us from assessing cognitive load impact.

Despite varying frequency of use, enactment of teaching activities was similar. The vast majority of teaching was oral. Although participants in our prior study[15] described physical teaching strategies such as hand gestures and manipulation of the equipment outside of the patient’s body, we rarely observed these; when they were used learners often did not notice them. Since procedural skills are inherently physical, it is interesting that more teaching was not enacted physically. This highlights a critical aspect of procedural teaching:

the teacher and learner cannot simultaneously have identical experiences, as may be more possible with cognitive tasks (e.g., patient interview, physical examination, or clinical reasoning). This is particularly challenging in procedures where only one person can control the equipment at a time, as is the case for many HPE procedures. In such cases, once the learner has established a baseline level of competence, she becomes the primary operator, while the teacher becomes the observer who must decide when and how to intervene. Experienced endoscopy teachers highlighted this as a challenge in our prior work[15], and the results of the present study illuminate approaches that teachers of single-operator procedures might use to optimize cognitive load and learning.

Because colonoscopy (like many other HPE procedures) entails substantial intrinsic and germane demands, our findings suggest that teachers should consider using teaching strategies judiciously, particularly when complex verbal teaching is needed. Colonoscopy is a complex activity, and most of our fellow participants were in their first year, having completed well below the 300-colonoscopy threshold generally required for competence[22,23]. Indeed, overall mental effort was high demonstrated by a median of 7 (1-9 scale) for Paas' method[18] with the lowest Paas score of 4 for the most experienced fellow. Continuously narrating during procedural teaching may overload the learner's working memory if it has reached capacity managing the intrinsic and germane load.

Timing of teaching

Our findings also suggest procedural teachers should consider timing of their teaching. When teachers taught during periods of intense cognitive effort, fellows felt unable to listen or respond to their instruction. Identifying a cognitively overloaded learner in real time could enable a teacher to modulate content, timing and intensity of teaching to reduce cognitive load, yet how teachers assess learner cognitive load has not been studied.

We observed that pre- and post-procedure periods were rarely used for teaching. These are times when learners' working memories should have greater capacity for listening to and processing teaching. In the cases where the teacher taught after the procedure, fellows perceived it promoted learning. The post-procedure period may also be an optimal time to deliver feedback, as the literature supports greater benefit for feedback given after, rather than during, during procedural learning[24,25].

Another useful time period for teaching was during takeover. Fellows reported that during takeover they were better able to listen; accordingly, the most common perceived cognitive load impact was increased germane load. This makes sense, as during takeover the fellow has only a single task – to learn – and had reduced from three sensory channels (visual, tactile and auditory) to two (visual and auditory). The "stop and listen" strategy (which we did not observe) might also free up working memory space for learners to hear instruction.

Familiarity with the learner

At the hospitals where we performed our observations, fellows are often paired with an attending who is unfamiliar with their level of competence. Our data suggest such lack of familiarity might affect learner cognitive load. When teachers asked questions that were too easy, or taught concepts fellows already knew, fellows perceived no cognitive load impact (in particular no benefit for germane load). Conversely, when the schemas or the clarifying questions surpassed the learner's level of competence, they were perceived as increasing extraneous load. These data imply that learners need teaching directed to their level of prior experience and understanding. A brief check-in with a learner to ascertain his level of competence is a simple way to assess competence before a procedural session begins. We observed this twice; in both cases, the fellows perceived benefit. A curricular approach would be scheduling teachers and learners to perform procedures together regularly over time, similar in concept to continuity clinics mandated for many graduate medical education programs, or longitudinal integrated clerkships in undergraduate medical education, which have proved beneficial for learning[26].

Autonomy

Although some teachers narrated frequently, four spent extended periods of time observing without speaking or intervening, granting the fellow more physical space. The four fellows all perceived this to promote germane load. This makes sense, as multi-channel processing can increase cognitive load[27]. Procedural tasks require two channels (visual and tactile), and the addition of the auditory channel may overwhelm working memory. Waiting for a period of less intense tactile and visual demands, or advising the learner to stop and pay attention to the teacher may increase cognitive capacity for verbal teaching.

There may be sociocultural benefits to providing autonomy as well. Many procedural learners may suffer from imposter syndrome[28], causing negative emotions that increase extraneous load[29]. Being empowered to function autonomously could reduce imposter syndrome and reduce negative emotions and extraneous load, thereby freeing working memory capacity for germane processing. Furthermore, autonomy is considered necessary for learners to develop intrinsic motivation[30]. Systems restraints may prevent fully entrusting learners to perform some procedures, yet limiting autonomy in procedural learning may reduce development of competence[31], whereas providing a developmentally appropriate degree of autonomy may motivate procedural learners[32].

Cognitive load types

The cognitive load impacts of each teaching activity were generally consistent with our predictions. We discuss each type of cognitive load below.

Germane load

Germane load was numerically higher than intrinsic or extraneous, and germane load was the type most often affected by teaching. This likely reflects that we observed mostly fellows in the first year (so they still have much to learn), yet that had some experience with colonoscopy (so the intrinsic load is not as overwhelming). Most of the teaching activities with perceived germane load benefit could promote self-regulated learning (e.g., feedback, schema teaching, checking understanding, expectation setting)[33]. Teaching self-regulation skills and enabling procedural learners to progressively take greater control over their learning may produce more mature and skilled proceduralists. Likewise, gradual reduction in intensity of teaching (i.e., granting more autonomy) as learners progress may promote self-regulation and enhanced procedural learning and reduce likelihood of expertise reversal effect[34]. This approach is aligned with competency-based training[35,36], yet may be difficult to implement if there is lack of continuity between teachers and learners, or in procedural settings where patients expect a fully trained expert to perform their procedure and where health systems demand efficiency and quality[37,38].

Whether germane load should be considered a unique construct, or whether it is best conceptualized as an aspect of intrinsic load is debated[6,7]. Intuitively, completing a task and learning from a task are related, yet require different cognitive activities. The moderate correlation between activities perceived to promote germane load and measured germane load provides one piece of evidence supporting germane load, and supports those teaching activities as useful for promoting procedural learning.

Intrinsic load

Teaching commonly affected intrinsic load. Although the primary goal of CLT is to match intrinsic load to the learner's level, we considered decreasing intrinsic load as a desirable outcome. This is because, for most of our participants, the overall task of completing a full colonoscopy was beyond their technical skills. To promote learning among relatively novice trainees such as ours, simplifying procedural tasks can help. The approaches predicted to reduce cognitive complexity (motor instruction) and psychomotor complexity (technical assistance, takeover) were perceived to decrease intrinsic load. Technical assistance is a part-task approach, and takeover provides a worked example, both of which can reduce intrinsic load among early learners[39]. Another way to reduce intrinsic load is to increase experience and competence, which could not be captured in individual observations. For many procedural skills, simulators are an effective means to increase skills in a risk-free environment[40,41]. Conversely, more experienced learners should perform full tasks to avoid the expertise reversal effect, where partial tasks or excessive teaching can hinder learning[39].

Extraneous load

Extraneous load was not prominent in our observations. We rarely observed potential sources of extraneous load, measured extraneous load was low, and relatively few teaching activities were perceived to affect extraneous load. Extraneous load may be limited in most routine ambulatory procedures like we observed, though procedures performed in more chaotic urgent/emergent settings would induce more extraneous load. Nevertheless, optimizing the environment to the extent possible in complex workplace settings may reduce extraneous load and promote mental space for learning[42].

Methodological considerations and limitations

Our findings should be considered within the context of our methodology. Considering our participants, only 25% were female, because the fellow classes from which we recruited had fewer females than average in our program, and nationally only about one-third of gastroenterology fellows are women[43]. While limited evidence in the social sciences suggests potential impact of gender on cognitive load[44], previous work did not find gender to have an impact on cognitive load in the endoscopy setting[12]. However, the present study provides no information as to how gender might affect teaching strategies attendings use.

Considering our observation methods, we performed the study within a single procedural field, but the teaching activities we studied are not specific to endoscopy, and appear useful to other procedures. Mindful of interactions between the observer, the participants and the research question, we intentionally chose to utilize a single observer. The observer was a gastroenterologist who was known and had good interpersonal relationships at the study sites (as recommended by Paradis and Sutkin[45]). The observing author has great familiarity with endoscopy (having performed several thousand endoscopic procedures previously), and was therefore most qualified to accurately capture behaviors and utterances. Furthermore, failing to capture occasional teaching activities would not likely impact the findings of the qualitative analysis which was the primary focus of this work. Limiting observation to one colonoscopy with each attending did not capture variation in teaching that might occur across teaching of different learners, but our interest was in variation among different teachers. Although our sample size of 10 observations was small, we reached information sufficiency of teaching activity enactment. For the quantitative analysis, given the small sample size[46], we performed only a single statistical test that focused on germane load.

Considering our measurement methods, we assessed fellow perceptions of how teaching strategies affected cognitive load, but the ultimate impact on actual cognitive load and learning is unknown. Studies suggest that learner perceptions of teaching and learning correlate poorly with actual learning[47], so our findings should be taken within this context. The correlation we found between activities perceived to promote germane load and measured germane load should also be interpreted with caution, since cognitive load measures are most valuable when coupled with measures of learning[48]. However, as our participants all had at least modest prior experience, we did not think it is possible to separate learning directly attributable to the single observed colonoscopy from prior learning. Additionally, we are not aware of any instrument that can quantify learning during colonoscopy; while performance scales exist[49], performance is a different construct from learning[50]. These considerations highlight a major challenge in performing research in actual (i.e., not simulated) HPE workplace. Finally, since we measured cognitive load and performed interviews after the colonoscopy was complete, the “peak and end effect” might occur (better recollection of recent or significant events). However, most participants recalled and reacted to teaching activities that occurred throughout the colonoscopies, suggesting limited impact in what were relatively short procedures.

CONCLUSION

CLT has clear relevance to procedural skills training in the health professions. The teaching activities that we studied were perceived to affect intrinsic, germane and extraneous load in mostly beneficial ways and were consistent with our predictions based on CLT. Our findings suggest that, to promote balanced cognitive load and learning, procedural teachers should be familiar with, and adapt their teaching to, their learners' prior experience and competence, and should be cognizant of their learners' cognitive state and modulate amount and type of verbal teaching, including periods of silence to allow learners to think and process, and promote autonomy. Similar studies performed in other procedural fields, and experimental studies in which cognitive load and performance are correlated with types and intensity of teaching activities, could provide additional validity evidence for our findings. We further suggest teachers should not miss opportunities to teach when learners have greater working memory capacity available, particularly before or after procedures. We conclude that awareness of CLT and use of CLT-informed teaching strategies has the potential to improve procedural teaching in the health professions.

REFERENCES

1. Wayne DB, Barsuk JH, McGaghie WC. Procedural training at a crossroads: striking a balance between education, patient safety, and quality. *J Hosp Med.* 2007;2(3):123-125.

2. Norris TE, Cullison SW, Fihn SD. Teaching procedural skills. *J Gen Intern Med.* 1997;12 Suppl 2:S64-70.

3. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci.* 1988;12:257-285.

4. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Med Teach.* 2014;36(5):371-384.

5. Kalyuga S. Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review.* 2011;23(1):1-19.

6. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. *Perspect Med Educ.* 2015;4(3):107-109.

7. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. *Perspect Med Educ.* 2015;4(3):119-127.

8. Sewell JL, Maggio LA, Ten Cate O, van Gog, T, Young JQ, O'Sullivan PS. Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions: BEME Guide No. 53. *Med Teach* 2018 Oct 17:1-15 [Epub ahead of print].

9. Kahi CJ, Imperiale TF, Juliar BE, Rex DK. Effect of screening colonoscopy on colorectal cancer incidence and mortality. *Clin Gastroenterol Hepatol.* 2009;7(7):770-775.

10. Kaltenbach T, Leung C, Wu K, Yan K, Friedland S, Soetikno R. Use of the colonoscope training model with the colonoscope 3D imaging probe improved trainee colonoscopy performance: a pilot study. *Dig Dis Sci.* 2011;56(5):1496-1502.

11. Sedlack RE, Coyle WJ. Assessment of competency in endoscopy: establishing and validating generalizable competency benchmarks for colonoscopy. *Gastrointest Endosc.* 2016;83(3):516-523.e511.

12. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. 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* 2017;92(11):1622-1631.

13. Sewell JL, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. *Med Educ.* 2016;50(6):682-692.

14. Coderre S, Anderson J, Rostom A, McLaughlin K. Training the endoscopy trainer: from general principles to specific concepts. *Can J Gastroenterol.* 2010;24(12):700-704.

15. Sewell JL, Bowen JL, ten Cate O, O'Sullivan PS, Shah B, Boscardin CK. Learning challenges, teaching strategies, and cognitive load: experience of seasoned endoscopy teachers. *Revisions Submitted to Academic Medicine.* 2019.

16. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. In: Thousand Oaks, CA, USA: SAGE Publications, Inc.; 2014.

17. Pourmand K, Sewell JL, Shah BJ. What Makes a Good Endoscopic Teacher: A Qualitative Analysis. *J Surg Educ.* 2018;75(5):1195-1199.

18. Paas F, Tuovinen JE, Tabbers H, VanGerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist.* 2003;38(1):63-71.

19. Flanagan JC. The critical incident technique. *Psychol Bull.* 1954;51(4):327-358.

20. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qual Health Res* 2015 Nov 27 [Epub ahead of print].

21. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ.* 2015;48(8):805-814.

22. Tassios PS, Ladas SD, Grammenos I, Demertzis K, Raptis SA. Acquisition of competence in colonoscopy: the learning curve of trainees. *Endoscopy.* 1999;31(9):702-706.

23. Spier BJ, Benson M, Pfau PR, Nelligan G, Lucey MR, Gaumnitz EA. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc.* 2010;71(2):319-324.

24. Dilly CK, Sewell JL. How to Give Feedback During Endoscopy Training. *Gastroenterology.* 2017;153(3):632-636.

25. Walsh CM, Ling SC, Wang CS, Carnahan H. Concurrent versus terminal feedback: it may be better to wait. *Acad Med.* 2009;84(10 Suppl):S54-57.

26. Teherani A, Irby DM, Loeser H. Outcomes of different clerkship models: longitudinal integrated,

hybrid, and block. *Acad Med.* 2013;88(1):35-43.

27. Mayer RE, Moreno R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist.* 2003;38(1):43-52.

28. Legassie J, Zibrowski EM, Goldszmidt MA. Measuring resident well-being: impostorism and burnout syndrome in residency. *J Gen Intern Med.* 2008;23(7):1090-1094.

29. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ.* 2012;46(11):1055-1062.

30. Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. *Med Teach.* 2011;33(12):961-973.

31. Dacey RG, Jr., Nasca TJ. Seniorization of Tasks in the Academic Medical Center: A Worrisome Trend. *J Am Coll Surg.* 2019;228(3):299-302.

32. Orsini C, Binnie VI, Wilson SL. Determinants and outcomes of motivation in health professions education: a systematic review based on self-determination theory. *J Educ Eval Health Prof.* 2016;13:19.

33. Panadero E. A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Front Psychol.* 2017;8:422.

34. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. *Educational Psychologist.* 2003;38(1):23-31.

35. Ten Cate O. Competency-Based Postgraduate Medical Education: Past, Present and Future. *GMS J Med Educ.* 2017;34(5):Doc69.

36. Carraccio C, Englander R, Van Melle E, et al. Advancing Competency-Based Medical Education: A Charter for Clinician-Educators. *Acad Med.* 2016;91(5):645-649.

37. Ten Cate O. Supervision and Entrustment in Clinical Training: Protecting Patients, Protecting Trainees. *AHRQ Web M&M.* 2018. Available online at https://psnet.ahrq.gov/webmm/case/461/Supervision-and-Entrustment-in-Clinical-Training-Protecting-Patients-Protecting-Trainees (accessed 15 April 2019).

38. Babbott S. Commentary: watching closely at a distance: key tensions in supervising resident physicians. *Acad Med.* 2010;85(9):1399-1400.

39. van Merrienboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010;44(1):85-93.

40. Singh S, Sedlack RE, Cook DA. Effects of Simulation-Based Training in Gastrointestinal Endoscopy: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol.* 2014;12(10):1611-1623.

41. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA.* 2011;306(9):978-988.

42. Lee MH, Schemmel AJ, Pooler BD, et al. Radiology Workflow Dynamics: How Workflow Patterns Impact Radiologist Perceptions of Workplace Satisfaction. *Acad Radiol.* 2017;24(4):483-487.

43. Accreditation Council for Graduate Medical Education. Data Resource Book, Academic Year 2016-17. Chicago, IL: Accreditation Council for Graduate Medical Education; 2017. http://www.acgme.org/About-Us/Publications-and-Resources/Graduate-Medical-Education-Data-Resource-Book. Accessed 15 April 2019.

44. Strombach T, Margittai Z, Gorczyca B, Kalenscher T. Gender-Specific Effects of Cognitive Load on Social Discounting. *PloS one.* 2016;11(10):e0165289-e0165289.

45. Paradis E, Sutkin G. Beyond a good story: from Hawthorne Effect to reactivity in health professions education research. *Med Educ.* 2017;51(1):31-39.

46. Leppink J, Winston K, O'Sullivan P. Statistical significance does not imply a real effect. *Perspect Med Educ.* 2016;5(2):122-124.

47. Sitzmann T, Ely K, Brown KG, Bauer KN. Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education.* 2010;9(2):169-191.

48. Leppink J. Cognitive load measures mainly have meaning when they are combined with learning outcome measures. *Med Educ.* 2016;50(9):979.

49. ASGE Training Committee, Sedlack RE, Coyle WJ, et al. ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD. *Gastrointest Endosc.* 2014;79(1):1-7.

50. Soderstrom NC, Bjork RA. Learning Versus Performance:An Integrative Review. *Perspect Psychol Sci.* 2015;10(2):176-199.

SUPPLEMENTAL MATERIAL FOR CHAPTER 6.

Qualtrics form used for coding.

What is the observation number? ____________

Please consider whether each Teaching Strategy below was discussed in the transcript. For each teaching strategy discussed in the transcript, code how fellow described impact on types of CL (selecting all that apply). For definitions of teaching strategies, refer to subsequent table.

	Strategy present?	Decrease IL?	Increase IL?	Decrease EL?	Increase EL?	Decrease GL?	Increase GL?
Check learner understanding	☐	☐	☐	☐	☐	☐	☐
Cognitive information	☐	☐	☐	☐	☐	☐	☐
Environment modifications	☐	☐	☐	☐	☐	☐	☐
Expectations/ roles	☐	☐	☐	☐	☐	☐	☐
Feedback/ praise	☐	☐	☐	☐	☐	☐	☐
Mastery mindset	☐	☐	☐	☐	☐	☐	☐
Motor instructions	☐	☐	☐	☐	☐	☐	☐
Stop and focus	☐	☐	☐	☐	☐	☐	☐
Takeover	☐	☐	☐	☐	☐	☐	☐
Teach schemas	☐	☐	☐	☐	☐	☐	☐
Technical assistance	☐	☐	☐	☐	☐	☐	☐
Other	☐	☐	☐	☐	☐	☐	☐

Definitions of strategies.

Strategies	Definitions
Check learner understanding	Attending asks questions to check fellow's understanding of what is happening during colonoscopy
Cognitive information	Attending makes statements to fellow that could alleviate fear, anxiety, worry, or other negative emotions
Environment modifications	Attending manipulates components of the environment (e.g., gurney, patient position, monitor position, fellow stance), or instructs other persons present to behave
Expectations/ roles	Attending articulates roles of themselves and/or the fellow, or tells the fellow what they can expect during entire colonoscopy or particular parts
Feedback/ praise	Attending tells fellow when they have done something good or bad, can be specific (i.e., feedback) or generic (e.g., "good" which would be considered praise)
Mastery mindset	Attending emphasizes good technique (i.e., rather than focus on independently completing colonoscopy)
Motor instructions	Attending tells fellow specifically what to do with colonoscope or instruments
Stop and focus	Attending instructs fellow to stop motor activity and listen
Takeover	Attending assumes primary control of colonoscope while fellow observes or assists with instruments (e.g., biopsy forceps)
Teach schemas	Attending explains what, how, or why (either specific to case or general approach/truisms)
Technical assistance	Attending performs part of task (e.g., biopsy forceps) while fellow maintains primary control of colonoscope

If "Other" teaching strategy was discussed in the transcript, describe here.

____

How (if at all) did fellow describe recognizing feeling of cognitive overload?

____

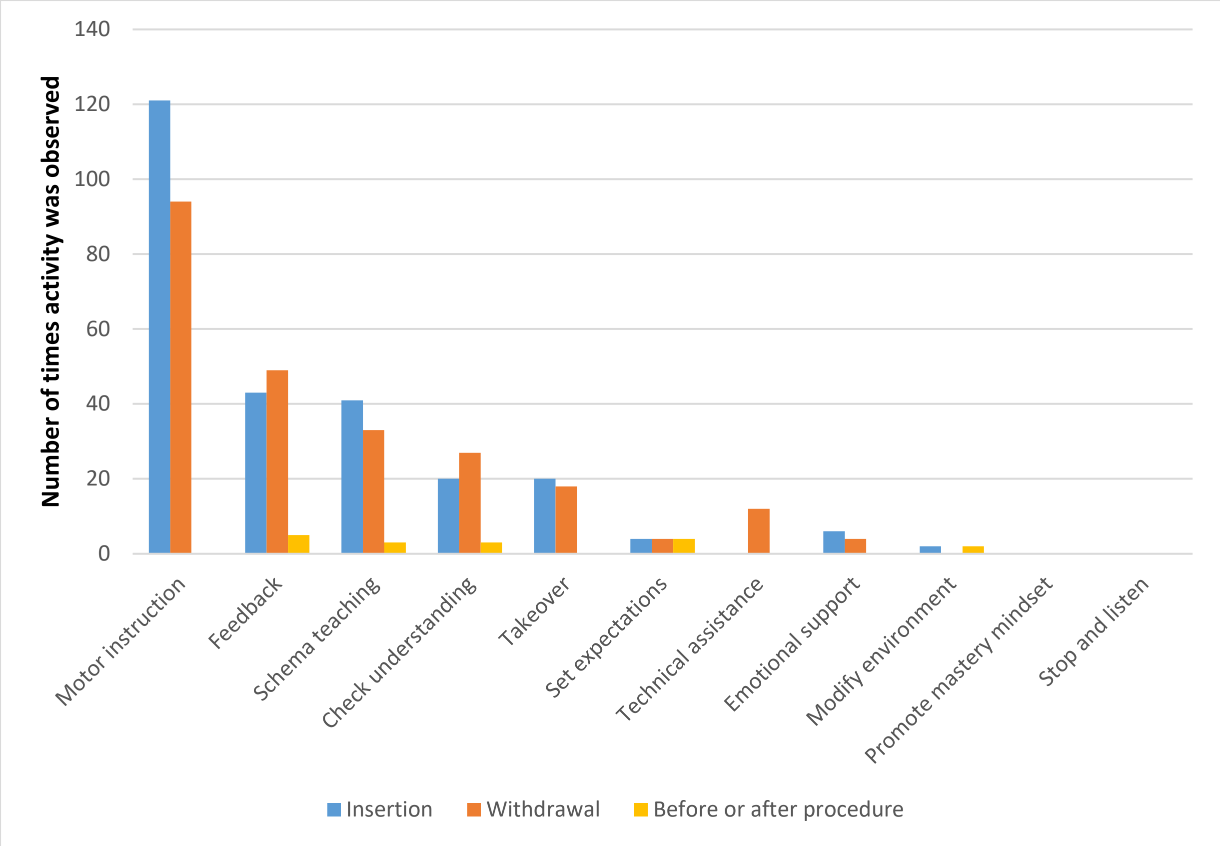
How (if at all) did fellow describe acting upon cognitive overload?

____

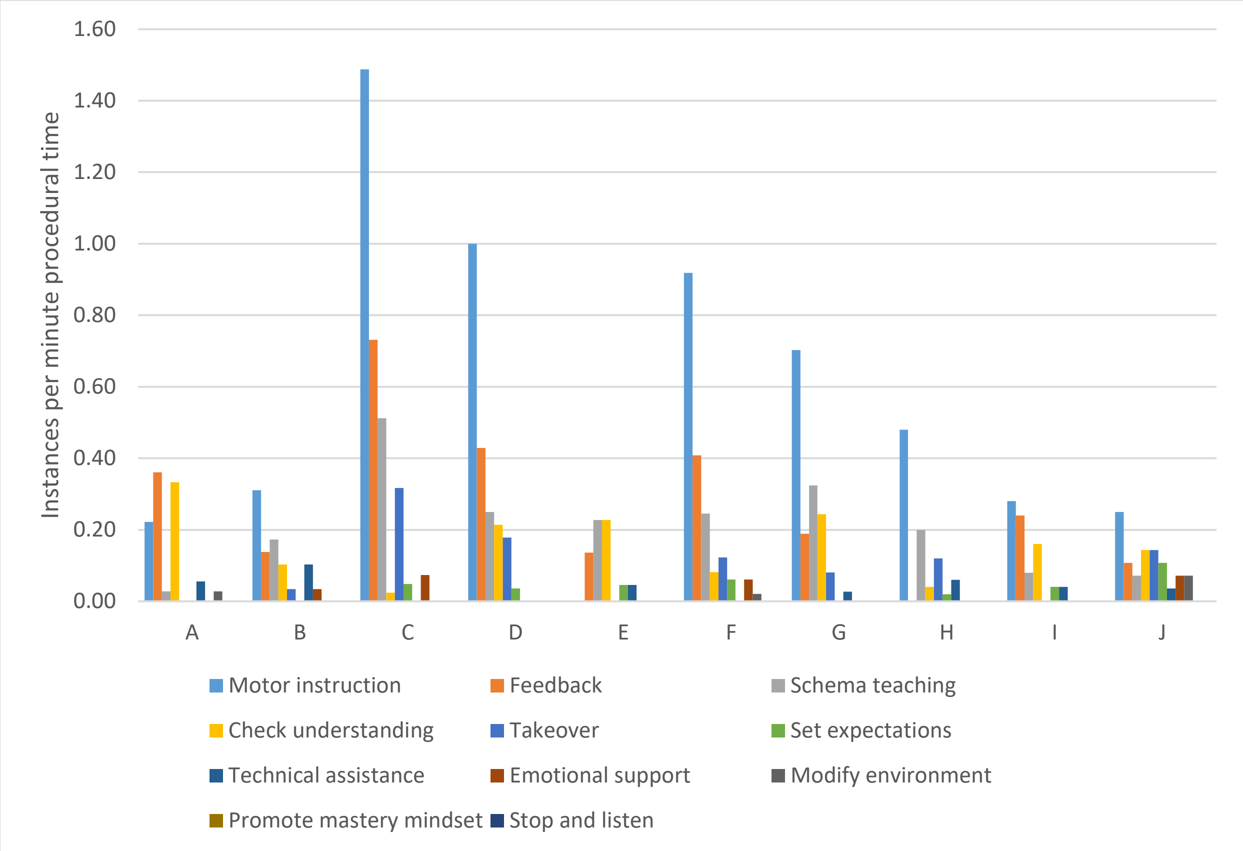
Please enter any additional comments.

____

Supplemental Figure 1. Teaching activities and associated portion of colonoscopy.

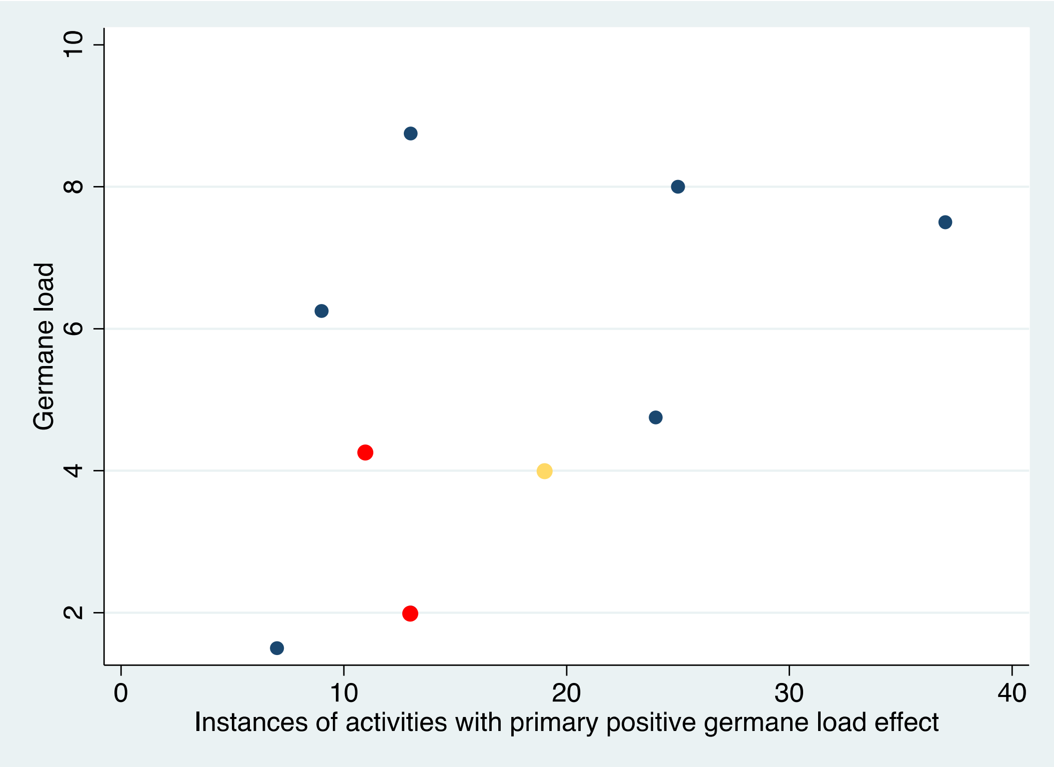


Supplemental Figure 2. Pattern of teaching activity use by attending.



A through J labels for attendings were randomly assigned and do not reflect actual order of observations.

Supplemental Figure 3. Graphical relationship between activities promoting germane load and measured germane load.



Blue points indicate 1st-year fellows. Yellow dot indicates two 1st-year fellows, both of whom had a measured germane load measured of 4 and for whom there were 19 instances of teaching activities that could promote germane load (though there were different patterns of which specific teaching activities were used). Red dots indicate 2nd-year fellows, who had significantly greater prior colonoscopy experience.

Chapter 7

Discussion

BACKGROUND

The work in this thesis supports the assertion that procedural skills training remains a major challenge for health professions educators. Unlike other primarily cognitive learning settings, procedural training requires learners to simultaneously learn and perform both cognitive and psychomotor tasks, each of which can be quite complex. Many health professions education (HPE) learners enter procedural training programs with little formal training in psychomotor tasks, and so the psychomotor learning domain presents a radical departure from the familiar task of learning and applying mostly cognitive information. Procedural settings exert other unique pressures on learners as well. Decisions during procedural tasks must be made quickly, yet can have immediate deleterious effects on patients. Learners in procedural fields may base much of their developing professional identity on their success performing procedures[1], yet many undergraduate HPE programs that focus mostly on cognitive learning may not prepare learners for the procedural learning they encounter in post-graduate residency and fellowship training[2]. Sociocultural aspects of many procedural specialties may exert additional pressures. Taken together, these factors contribute to high risk of cognitive overload in early procedural learners, as reflected in the story that began this thesis.

We are at a critical juncture in procedural training in the health professions. While procedural complexity, sophistication, and sub-specialization in both medical and surgical fields continues to advance[3,4], both undergraduate medical education (UME) and graduate medical education (GME) programs are increasingly restricted and regulated[5,6], and the public demands increasing levels of accountability and transparency for assessing and reporting on competence of health professionals[7]. Consequently, there is an urgent need to better understand how trainees learn to perform procedural skills, and how teachers and training programs can optimize training and learning.

The studies described in this thesis all focus on this need, by examining how cognitive load theory (CLT) can inform effective approaches to teaching and learning procedural skills. We selected CLT as the primary theoretical framework because cognitive processes are central to all aspects of procedural learning (even psychomotor tasks are stored as cognitive schema in long-term memory), yet working memory capacity for procedural learners remains limited. Other learning paradigms such as behaviorism and sociocultural learning are undoubtedly relevant to procedural learning, yet accurate schemas for cognitive and psychomotor concepts must ultimately be created, refined, and accessed for a learner to become a competent proceduralist. We therefore consider CLT as the learning paradigm positioned most closely to actual learning, and we believe that other influences (such as sociocultural influences) likely exert influence, at least in part, through cognitive load[8].

CLT builds upon the Atkinson-Shriffin model of human learning[9] by characterizing the different ways that the working memory can be used, namely completing essential task elements (intrinsic load), forming and refining learning schema (germane load), and attending to distractions or other non-essential information (extraneous load)[10,11]. Considering the learning task, the learner, the setting, and the teaching approach should enable a procedural teacher to promote goals of CLT in their learner: to match intrinsic load to the learner's level of competence, to optimize germane load, and to minimize extraneous load. CLT is relevant to HPE and is increasingly studied, yet how and when CLT tenets should be applied within authentic HPE workplaces, including procedural skills training, remains understudied and inadequately characterized. For additional information about CLT, we refer the reader to the detailed description of CLT provided in the Introduction.

SUMMARY OF FINDINGS

The research reported in this thesis seeks to better understand how CLT informs procedural teaching and learning in the health professions. Procedures performed across the health professions have striking heterogeneity in terms of duration, complexity, risk and learning curves. Therefore, to develop a cohesive body of research, we selected gastrointestinal (GI) endoscopy as an exemplar procedural setting that we considered "middle-of-the-road" in terms of complexity, duration and learnability. We designed our studies to address five primary research questions, that, while performed in the setting of GI endoscopy, would produce implications that would be generalizable to a wide variety of procedural settings. These were described in the Introduction and are listed again below:

1) How can we measure intrinsic, extraneous and germane cognitive load among colonoscopy learners?
2) How do features of learners, tasks, settings and teachers influence intrinsic, extraneous and germane cognitive load during colonoscopy training?
3) How has CLT been used to study teaching and learning within professional workplace settings, both within and outside of the health professions?
4) What learning challenges do experienced GI endoscopy teachers see their learners experience, what teaching strategies do they use to help learners overcome these challenges, and how do those strategies align with cognitive load theory?
5) How do "everyday" teachers of GI endoscopy teach, and how does their teaching affect learners' perceived intrinsic, extraneous and germane cognitive load?

Each of these research questions is addressed by a study, the major findings of which are summarized below. This brief summary will serve as a prelude to discussing broad themes address by this thesis, directions for future research, and strengths and limitations of our research. For a more detailed summary of each study's findings, please see the Summary section, which follows the Discussion.

Chapters 2 and 3 examine methods to measure cognitive load, and explore factors that may impact cognitive load, during the exemplar procedural training setting of colonoscopy. Chapter 2 documents a novel investigator-developed self-report instrument that separately estimated intrinsic, extraneous and germane cognitive load among nearly 500 colonoscopy learners across the United States. The chapter presents strong evidence for validity, including exploratory and confirmatory factor analysis yielding satisfactory fit parameters, and supports a three-factor model for cognitive load, corresponding to intrinsic, extraneous and germane load. Chapter 3 considers a wide array of factors related to the major contributors to cognitive load in procedural environments: the learner, patient/task, setting and supervisor, and models unique sets of factors associated with each type of cognitive load. This information may be used by procedural teachers to target specific cognitive load goals through direct teaching and curricular design. Chapter 4 marks a transition from understanding the role of CLT, to examining how to teach procedures through the lens of CLT. In this chapter, we examine the literature to understand how others have used CLT to study professional workplace teaching and learning. Chapter 4 offers the broadest-reaching systematic review of CLT in HPE available in the literature, systematically synthesizing a vast amount of information from 116 studies performed across multiple professions (both within and outside of healthcare) to offer multiple theoretical and practical implications. Finally, we apply lessons learned in Chapter 4 to teachers and learners of procedural skills, returning to the realm of gastrointestinal endoscopy. Chapter 5 characterizes how experienced GI endoscopy teachers see procedural learners struggle, and how they approach procedural teaching to help learners overcome those challenges. The teaching strategies they describe relate clearly to goals of CLT. Chapter 6 characterizes how 'everyday' endoscopy teacher enact various teaching activities while being observed during actual teaching of endoscopy learners in real colonoscopy cases. The chapter further analyzes how the learners perceive those teaching activities as affecting intrinsic, germane and extraneous cognitive load.

MAJOR CONTRIBUTIONS

While individual chapters address the five research questions described above, the body of work as a whole furthers understanding of procedural skills teaching and learning, and CLT itself, by addressing several major questions, listed below. The multiple chapters that address each of these questions are indicated in parentheses. The first two topics are more theoretically-oriented, while the second two address more practical issues. Each of these topics will be considered in turn, linking research documented throughout this thesis with other literature.

1) How can we measure cognitive load subtypes (intrinsic, extraneous and germane load) in procedural skills training settings? *(Chapter 2, 3, 4 and 6)*
2) Should germane load be treated as a unique subtype of cognitive load, or as an aspect of intrinsic load? *(Chapters 2, 3 and 6)*
3) What contributes to intrinsic, germane and extraneous load in procedural skills training? *(Chapters 2, 3, 4, 5 and 6)*

4) How can we leverage teaching and curricular innovations to promote balanced cognitive load and promote learning during procedural skills training? *(Chapters 4, 5 and 6)*

Measuring cognitive load

The ability to measure or estimate cognitive load is critical for research, both to help the theory evolve, and to further understanding of how it should be used to inform curriculum design, workplace environments and direct teaching. Yet, measurement of cognitive load remains a critical challenge that is in many ways a roadblock to furtherance of CLT[12-16]. There are several constructs related to cognitive load that researchers have attempted to measure or estimate. In Chapter 4 we found that some of these terms (the first three in particular) were often used interchangeably across different studies, yet CLT scholars appreciate subtle differences among them. It will be helpful for the reader to be familiar with these terms, as noted below.

- *Cognitive load* is the working memory demand imposed while performing a task[11].
- *Mental load (or mental workload)* is the aspect of cognitive load that arises from the interplay between learner capability and task characteristics/complexity[17].
- *Mental effort* is the portion of cognitive capacity that the learner actually allocates to performing a task[17].
- *Intrinsic load* is the subtype of cognitive load that occurs as learners accomplish essential task elements[11] (also considered "productive" cognitive load[18]).
- *Extraneous load* is the subtype of cognitive load that occurs when learners attend to non-essential internal or external stimuli[11] (also considered "unproductive" cognitive load[18]).
- *Germane load* is the subtype of cognitive load that occurs when learners form or refine learning schema[11].

All of the above constructs are challenging to measure (estimate might be a more appropriate term), because they are invisible processes that occur within the working memory of the learner. Multiple approaches have been used to estimate cognitive load[11,12], but all seek to quantify complicated constructs that are not immediately intuitive. This begs the question: *Are we measuring what we intend to measure when we study these cognitive load constructs?* The work in this thesis makes significant contributions to understanding how to measure or estimate cognitive load constructs, in particular how to measure cognitive load subtypes, and whether the constructs being measured are actually the cognitive load types.

By providing a detailed review of the world's literature regarding cognitive load in professional workplace education, Chapter 4 helps us understand how scholars are measuring cognitive load related constructs. Of the 116 studies reviewed, nearly 90% sought to estimate cognitive load-related constructs, using a variety of approaches, including psychometric instruments, secondary task approaches (response time and accuracy), physiologic measures and qualitative methods. A handful of studies compared measurement approaches, but they did not provide adequate evidence to identify a "best approach". Rather, each method of estimating cognitive load has pros and cons. We next discuss each measurement approach, including how work in this thesis informs them.

Psychometric instruments

Psychometric instruments are designed to gain insight into a person's thoughts through the use of printed words and numeric scales. Psychometric instruments are easy to use, cost little, require little to no technology, and interfere minimally with learning tasks (when used after the task is complete, as in most studies), so it is not surprising that they were by far the most common approach taken in studies reviewed in Chapter 4. However, psychometric means to estimate cognitive load constructs have several potential shortcomings that may negatively impact validity of their measurements.

Most studies in Chapter 4 utilized existing scales such as Paas' single item[17] or the NASA-Task Load Index (NASA-TLX)[19], yet few provided any evidence for validity of those existing scales within their particular setting. While the Paas scale is simple and easily understood by learners it provides only a single estimate of overall mental effort. The NASA-TLX provides more granular data, yet is more complex and might be less readily applicable to diverse settings (negatively affecting content evidence for validity), or might not be fully understood by participants (negatively affecting response process). Of the handful of studies that developed a new or substantially revised instrument, very few provided rigorous evidence for validity as recommended by Artino et al[20].

Conversely, we fully followed the guidelines outlined by Artino et al to develop the CLIC and present a detailed description of evidence for validity (Chapter 2). Developing the CLIC took well over a year, yet this time was rewarded with an exploratory factor analysis showing a clear three-factor structure corresponding to intrinsic, extraneous and germane load with all items fitting into their expected factor, and confirmatory factor analysis that demonstrated excellent fit parameters. Chapter 3 provides additional evidence that the CLIC measures three separate constructs, as unique sets of features were associated with each cognitive load subtype. Furthermore, the features contributing to each cognitive load subtype made sense: features of learners and patients/tasks were associated with intrinsic load, features of the setting and supervisor (plus learner fatigue) were associated with extraneous load, while germane load was associated with the supervisor. Chapter 6 provides a final piece of data supporting a strong validity argument for the CLIC. Despite small numbers, germane load as measured by the CLIC showed a moderate correlation with teaching activities that fellows reported as beneficial for their learning. Although other investigators have developed psychometric instruments to estimate cognitive load subtypes in HPE workplace settings[21-30], none provided such rich validity data as the CLIC. The above considerations permit us to conclude that the CLIC measures three separate constructs, which we believe correspond to intrinsic, extraneous and germane load. Lacking a "gold standard" measure with which to compare, we cannot prove these to be the three cognitive load subtypes, yet we believe the evidence supports this as likely. We therefore propose that a carefully and rigorously developed psychometric instrument can be used to separately estimate cognitive load subtypes. However, there are additional threats to use of psychometric instruments in procedural skills research.

Any instrument intended to measure cognitive load subtypes will include multiple items that take time and thought to answer. As opposed to cognitive tasks, which might be more "interruptible," pausing in the middle of an actual medical procedure for the learner to complete a series of questions is nearly impossible. Consequently, we asked fellows to complete the CLIC after they had finished the procedure, an approach that was nearly universal in the studies in Chapter 4 that occurred in actual workplaces. This means that CLIC and other 'post-hoc' instrument scores represent an impression of the overall amount of cognitive load that occurred over the course of a procedure. Such post-hoc measures may be less accurate due to forgetting or bias related to the learner's performance (that is, a learner who performed well may perceive post-hoc cognitive load differently than one who struggled), and necessarily miss granular detail regarding variability in cognitive load throughout a procedure, which was reported in a few studies in Chapter 4[28,31,32].

Another challenge is that the amount of time and effort needed to develop and test a new psychometric is prohibitive for many researchers. The completed instrument is a static entity, and if emerging evidence suggests a need to alter the instrument, the need to collect validity evidence recurs. Furthermore, an instrument such as the CLIC (Chapter 2), developed for a single type of procedure, is applicable only to that single procedure, limiting overall utility. Nevertheless, considering our success developing the CLIC by adapting from an instrument designed for an entirely different setting (statistics teaching of graduate students[33]), we believe that the CLIC provides a model that could be further adapted for use in other procedural settings, and potentially even cognitive learning settings as well.

How might psychometric instruments be more readily applied to CLT research in procedural training settings? Single-item ratings of cognitive load might address the post-hoc concerns, as they are brief and can be more easily used at various time points throughout an actual procedure, yet it is difficult to know what to do with these data, as they are a summation of multiple subtypes of cognitive load, which have varying impact on learning. Alternatively, simulated procedures are more easily interrupted (since actual patient care is not occurring), and so multi-item scales could be used more frequently. Indeed, 78% of studies in Chapter 4 were of simulation. However, we assert that, for procedural skills, most simulated environments lack adequate authenticity to fully represent the dynamics of the various cognitive load subtypes that occur in authentic procedural settings. Although simulation technology continues to improve, Chapters 5 and 6 suggest that many procedural learners have significant anxiety related to performance of procedures on actual patients, and it is not at all clear whether CLT research in simulated procedural environments will translate adequately to actual procedures performed on real patients.

Lacking a "gold standard" against which to compare psychometrically measured cognitive load constructs, we rely on research reporting evidence for validity so that we can circumstantially estimate the potential validity of each instrument. To address this and the foregoing challenges, we see value in developing a more generalizable psychometric instrument that can estimate cognitive load subtypes for a wide variety of HPE procedures. There is precedent for such an approach in the NASA-TLX[19] which is widely used across numerous fields within and outside the health professions (and was used in 46 studies in Chapter 4) and has been adapted for specific use in surgery (the SURG-TLX[34]). Though the NASA-TLX is often used in CLT research, it does not estimate the cognitive load subtypes, which we consider critical for CLT research to produce findings that are practically useful in procedural training settings.

Despite the shortcomings of psychometric approaches, they currently remain the only method to estimate the cognitive load subtypes and hence will be relevant for the foreseeable future.

Secondary task approaches

When a learner's working memory is full or nearly full, he will have diminished ability to attend to stimuli that are not critical for the task at hand (i.e., secondary tasks). This can be leveraged to provide an indirect estimate of overall cognitive load by asking learners to respond to a secondary stimulus, typically auditory, visual, or vibrotactile. A learner who is more cognitively loaded will respond less accurately and with greater delay. This approach was used by 24 of the studies reviewed in Chapter 4 (many of which were in procedural environments) and results generally were as predicted: less experienced learners and more complex tasks were associated with less rapid and less accurate responses to the secondary stimuli. This approach addresses some of the shortcomings of psychometric approaches by interfering less with procedures (the equipment is typically unobtrusive) and can therefore be used to assess moment-to-moment cognitive load (as in several studies reviewed in Chapter 4). Other than equipment costs, the main downside is that secondary task approaches only measure overall cognitive load (or mental effort), not cognitive load subtypes, which significantly limits utility of data.

Physiologic measures

A number of physiologic measures may be used to estimate cognitive load; 19 studies reviewed in Chapter 4 used physiologic measures such as heart rate variability, pupil dilation, skin conductance, and electroencephalograms. Like secondary task approaches, many physiologic measures provide moment-to-moment data, yet equipment may be costlier and more obtrusive in the procedural environment. Functional MRI (fMRI)[35] may be considered a physiologic measure and has potential to provide rich information that might imply cognitive load subtypes, yet it is very costly, can only be used with simulated tasks, and is challenging to use for many procedural tasks given physical constraints of the fMRI equipment. However, efforts exist to develop fMRI-compatible equipment to simulate more complex tasks like laparoscopic surgery[36].

Qualitative measures

Only one study reviewed in Chapter 4 sought to assess cognitive load qualitatively[37]. Although CLT constructs are intuitively quantifiable, qualitative approaches may address threats to validity inherent in quantitative instruments, particularly related to response process and the challenges of quantifying abstract processes of the working memory. We used qualitative methods in Chapter 6, not to measure cognitive load, but to determine directionality of impact that teaching activities had on procedural learners' cognitive load. We used simple terminology to ask these questions, asking how teaching helped the learner complete the task (intrinsic load), learn from the task (germane load), and focus or minimize distractions during the task (extraneous load). We found it straightforward to classify participants' responses and had excellent agreement amongst coders. Evidence for validity of the approach was provided by correlating perceptions of germane load impact with measured germane load. The small sample size prevented us from performing further confirmatory statistical tests, but we believe Chapter 6 supports a role for qualitatively assessing cognitive load and its subtypes in future research.

Centrality of cognitive load subtypes

This thesis focuses throughout on subtypes of cognitive load. Like others[38,39], we consider the ability to measure subtypes, rather than simply overall cognitive load, as critically important for furtherance of CLT and to provide practical information for designing targeted interventions to promote balanced cognitive load in procedural training settings. Overall cognitive load (or mental effort or mental workload) are crude measures, considering the diverse activities that the working memory can engage in, and the divergent impacts of each. It does not make sense to lump together three different constructs with very different impacts: germane load promotes learning, while extraneous load interferes with learning, and intrinsic load can be either beneficial for learning (when matched to the learner's level of experience) or detrimental (if too high or too low). Understanding how an element of a learning environment affects one of the cognitive load subtypes will much more effectively help the procedural teacher develop and test a potential solution than simply knowing how overall cognitive load is affected. Despite the importance of cognitive load subtypes, only eight of the studies in Chapter 4 attempted to measure them (and two of those eight studies are reports of the research described in Chapters 2 and 3 of this thesis). Furthermore, less than one-quarter of studies even mentioned any of the cognitive load subtypes! This reflects that most of the world's literature regarding cognitive load in workplace learning has tended to leverage the concept of cognitive load narrowly as an outcome to compare two or more educational conditions, rather than as a theoretical lens to promote understanding of how to teach and learn in these complex environments.

This thesis provides empirical data supporting the relevance of cognitive load subtypes to procedural learning. Chapter 2 supports separate identities for the three constructs we measured. If we had measured only overall cognitive load in Chapter 3, we would not know how to interpret the resulting model. In Chapter 5 we found that teachers of GI endoscopy use a wide variety of teaching strategies, each of which had clear relevance for one of the cognitive load types. This guidance helps procedural teachers know when to use particular teaching strategies. If a learner is struggling to complete the essential task (i.e., intrinsic load is very high), a teaching strategy aimed at reducing intrinsic load is appropriate, not a strategy to reduce extraneous load. If a learner doesn't seem to be "putting the pieces together" (i.e., germane load is low) during procedural learning, asking clarifying questions (which promotes germane load) might be more useful than physically assisting with the procedure (which reduces intrinsic load). Such differential effects were confirmed in Chapter 6, where fellows perceived divergent effects of various teaching activities on cognitive load subtypes. Clearly, the tenets of CLT are based on its subtypes and should be of primary focus in CLT-related HPE research.

As noted above, though, measuring cognitive load subtypes is more challenging than measuring overall cognitive load. When then should we consider overall cognitive load? Overall cognitive load may be a useful outcome to measure in experimental research, wherein a single change is made, for example to the learner's or teacher's approach, to the task, or to the learning environment. If only a single factor is modified, and the likely cognitive load subtype impact is predicted, then differences in overall cognitive load between conditions can be inferred to relate to that particular subtype. This was the case for a number of studies reviewed in Chapter 4, in which the inferred cognitive load subtype was often evident. Another potential use of measuring overall cognitive load is to identify learners who are cognitively overloaded. This is particularly important for procedural environments, where a mistake made by an overloaded learner could cause immediate physical harm to the patient. In such a case, the source of cognitive overload is not immediately pressing; the procedural supervisor simply needs to recognize the overloaded state and intervene.

What is germane load?

Germane cognitive load is not part of CLT's earliest descriptions[10], but is a more recent addition, based on the hypothesis that simply reducing extraneous load would not automatically lead to learning; rather, learners must engage in activities that promote schema formation and revision[40,41]. Since its introduction, scholars have debated whether germane load should be considered a unique form of cognitive load, or as an aspect of intrinsic load[13,42,43]. Leppink in particular has argued strongly for the latter conceptualization[43-45]. Because germane load includes activities that promote learning, the best evidence for its uniqueness would be strong correlation with learning (or performance as a proxy for learning), as suggested Naismith and Cavalcanti in a letter to the editor responding to the research presented in Chapter 2[46]. Unfortunately, few studies have examined associations between germane load and learning or performance. More than half of Leppink's experiments identified three-factor solutions for cognitive load, yet measured germane load was not associated with measures of learning[33,47]. Debue et al studied hypermedia learning, finding inverse association between germane and extraneous load, no association between germane and intrinsic load, and positive association between germane load and motivation; germane load was not associated with performance[48]. In both Leppink's and Debue's work, little time passed between measurement of germane load and assessment of learning. It takes time for schemas to be formed and stored in long-term memory (hence the benefit of spaced learning over massed learning[49]), and so the short interval might account for lack of correlation between germane load and learning. Alternatively, the high cognitive load of the learning activity in those studies could have led to a period of "working memory resource depletion," which was shown to manifest as a decrement in performance immediately following a prolonged period of mental effort in an experimental primary school setting of massed versus spaced learning[50]; this concept was discussed in a recent narrative review of CLT by Sweller, van Merriënboer and Paas[51]. Either way, we suspect greater delay is needed between measurement of germane load and assessment of performance.

Relationship to other variables is only one piece of validity evidence among many, which we argue should not be considered superior to other forms of validity evidence in a modern conceptualization of validity[52,53], and, furthermore, learning is only one other variable of relevance. Chapter 2 provides strong evidence for validity of the CLIC, which clearly identified a three-factor structure consistent with separate identities for intrinsic, extraneous and germane cognitive load. Chapter 3 provides additional evidence for three types of cognitive load based on relationships to numerous variables other than learning. Chapter 4 reviews six studies that examined germane load (two of which were published versions of Chapter 2 and 3). Of the four others, three supported germane load as a separate construct[23,30,54], while one did not[29]. Chapter 6 shows correlation between activities expected to promote germane cognitive load and measured germane load.

Because all of our research was performed in, and focused on, workplace settings (particularly procedural skills settings), it would be incredibly challenging to show a clear correlation between germane load and performance. Procedural learners rapidly gain skills and knowledge from numerous activities across training programs, and so correlating learning or performance with a single or even several measurements of germane load among such learners is illogical. Experiments among true novices (for example, Leppink's research[47]) address this issue, yet translation from simulation among novices to actual procedures performed by somewhat experienced learners may be limited. When planning this thesis research, we originally considered an experimental study design using an endoscopy simulator with novices. However, we were concerned that translation to authentic complex workplace procedural settings might be limited, and so we chose instead to perform the research described in Chapter 6, the results of which we feel are much richer and relevant to actual procedural skills training in the health professions than a study of simulation would have been.

We conclude that more data support than refute germane load as a separate type of cognitive load, and we appreciate the clear content validity of its separateness[13]. Intuitively, a motivated learner who uses strategies to promote schema formation should learn more (i.e., have greater germane load) when performing a task when compared with a learner who has identical prior experience yet who lacks motivation to learn and/or skills to promote schema formation. Despite lack of strong correlations between measured germane load and learning, a great deal of other evidence points to separateness, as described above. Of equal importance is that there exist teaching strategies that can be used to promote germane load independent of intrinsic or extraneous load, both generally[55,56] and in procedural settings (Chapters 5 and 6).

Leppink proposed that germane load is better conceptualized as "working memory resources allocated to dealing with intrinsic load"[44]. We agree that germane load does indeed represent working memory resources allocated to dealing with intrinsic load, as this another way to describe schema formation and revision. However, we argue that this complex label may downplay the vital importance of germane load for learning, and that there are ways to promote germane load other than optimizing or titrating intrinsic load. Accordingly, we recommended continued use of the term "germane load," recognizing its close relationship with intrinsic load. Despite its importance, germane load remains understudied, as demonstrated in Chapter 4, where only 19% of the 116 reviewed studies even mentioned the term "germane load", and we see great importance of dedicated study of germane load, particularly within procedural skills settings and other workplace settings in HPE, given the complexity of learning in these settings.

Contributors to cognitive load in procedural skills training

Chapter 4 confirms that relatively little research has applied CLT broadly to study procedural skills training in HPE – though multiple studies of procedural HPE settings were identified, most leveraged the concept of cognitive load, mental effort or mental workload superficially as a single outcome for comparing two or more procedural training conditions. Those studies promoted little understanding of factors that could impact different types of cognitive load, and did not substantially inform how CLT might be leveraged to improve procedural skills teaching and learning. The original research in this thesis contributes to these two important topics. In this section we discuss contributors to each cognitive load type and in the next section we address practical considerations for optimizing cognitive load among procedural learners.

Intrinsic load

Our research supports a traditional understanding of intrinsic load, which arises from completing the essential elements of a task. Chapters 2 and 3, plus the vast majority of studies reviewed in Chapter 4, strongly support that intrinsic load will be higher when tasks are more complex and/or learners have less experience, familiarity and knowledge; this holds true for procedural settings. Studies in Chapter 4 nearly universally found higher intrinsic load in complete versus partial tasks, actual versus simulated tasks, unfamiliar versus unfamiliar tasks, and less experienced versus more experienced learners. These strong and intuitive associations are incontrovertible, and we conclude that no further study of these links is warranted.

Our research does shed novel light on specific task elements that may contribute to intrinsic load in procedural skills settings. Each procedure in HPE includes a series of partial tasks, both cognitive and psychomotor, that must be completed in sequence and/or simultaneously to accomplish the full task. When designing intrinsic load items in the CLIC (Chapter 2), we deconstructed these tasks in consultation with other endoscopy experts, resulting in a roughly equal split between cognitive and psychomotor tasks, which exhibited similar factor loadings. This suggests that cognitive tasks may contribute to intrinsic load of procedural skills as much as psychomotor tasks. Chapter 5 suggests teachers recognize the importance of both, as many discussed learner challenges and teaching strategies related to both psychomotor and cognitive tasks. However, both Chapters 5 and 6 suggest learners may be more focused on the psychomotor aspects.

A combined cognitive and psychomotor task that was perceived as a major contributor to task complexity by teachers (Chapter 5) and learners (Chapter 6) alike was understanding how the colonoscope should feel, and interpreting that haptic and tactile feedback to generate a mental model of what was happening inside of the patient. This is a particular challenge for GI endoscopy procedures because the instruments are long (up to 2 meters) and flexible and can take on very complex looped configurations within the patient[57]. At the same time, the endoscopist's senses are limited to the cross-sectional two-dimensional video information (visual) and the feel of the instrument (tactile) to use as data for developing a mental model of instrument configuration. Similar challenges occur in a wide array of procedural settings, such as laparoscopic surgery, endovascular interventions (like cardiac catheterization), imaging-guided biopsy, lumbar puncture, or ultrasound-assisted placement of a central venous catheter. Each of these procedures (and countless others) involves limitations in sight and/or touch, increasing learners' challenges to develop mental models of what is happening inside the patient whose bodily defenses they are invading.

The association between fatigue and intrinsic load (Chapter 3) was a novel finding with broad implications for systems considerations, curricular design and direct teaching. While fatigue does not alter the actual task complexity, it may alter a learner's perception of complexity by diminishing the ease of retrieving existing schema from working memory. Although we did not assess performance, fatigue is known to negatively affect HPE trainee performance in clinical settings[58]. Another unique finding was that year in training and prior colonoscopy experience were independently inversely associated with intrinsic load. This suggests there are procedural learning opportunities outside of the procedural suite. While simulation provides well-described procedural learning benefits[59], there are likely numerous other opportunities, both cognitive (e.g., lectures, case discussions, observing others perform procedures) and sociocultural (i.e., as a learner becomes enmeshed in the procedural culture, they may develop and refine procedural schemas).

Extraneous load

Extraneous load occurs whenever learners utilize working memory resources to think about or focus on anything that does not promote task completion or learning. Studies reviewed in Chapter 4 investigated how data presentation, task standardization, multi-tasking, time pressure, and emotion could contribute to extraneous load. Though most of those studies were not in procedural settings, these sources of extraneous load are probably relevant to procedural learning.

Our research suggests that extraneous load is relatively minimal in GI endoscopy suites; in both Chapters 2 and 6, extraneous load as measured by the CLIC was numerically lower (1.8 in Chapter 2, 1.6 in Chapter 6) than intrinsic (3.1 and 3.0, respectively) or germane load (5.0 and 4.5, respectively). The low extraneous load scores likely reflect that the CLIC focuses on external sources of distraction; we included only one internal distraction item in the CLIC which did not meet inclusion criteria during factor analysis. External distractions may be minimized in the GI endoscopy suite, which most often involves elective, ambulatory procedures lacking the urgency or chaotic nature of an emergency procedure. Indeed, teachers (Chapter 5) and learners (Chapter 6) alike felt there was not much extraneous load from external sources. Teachers (Chapter 5) suggested endoscopy learners might have more internal sources of extraneous load, such as negative emotions, low self-efficacy, worries about competing demands (i.e., consult service patients to be seen), and stress related to the relationship between the teacher and the learner, but these were rarely mentioned by learners (Chapter 6). Internal sources of extraneous load in procedural settings warrant further study.

Teachers have long been considered a potential source of extraneous load. Classic instances of extraneous load have been described in classroom settings, such as when the lecturer's slide content does not match what he is talking about, or when figures or animations are excessively complex or distracting[60]. In such cases, students engage in unproductive cognitive processing to try and disentangle the mismatch between their auditory and visual channels; this limits their ability to listen to the lecturer and form/refine learning schemas. Our research confirms that teachers can induce extraneous load, and identifies both familiar and novel types of teacher-induced extraneous load. A familiar source is unclear or confusing instructions, which are essential elements of the CLIC (Chapter 2) and are not surprising. What is novel is that when learners in Chapter 3 perceived their teacher to have low confidence or to be less engaged, they had increased levels of extraneous load. This reflects the high stakes of HPE procedures and the great dependence of the learner on the teacher. A learner's lack of faith in their teacher's procedural skills or teaching may be a source of distraction or negative emotions that could contribute to extraneous load.

Chapter 6 reveals a novel source of teacher-induced extraneous load that was not identified in the studies reviewed in Chapter 4. Several participants in Chapter 6 reported the amount and intensity of the teacher's instruction contributed to extraneous load. In these cases, the teacher narrated almost continuously during the observed colonoscopy, using up to 3.3 teaching strategies per minute (more than one every 20 seconds!). When procedural learners have already reached full working memory capacity managing the essential task demands, even "good" teaching can adversely affect learning by inducing extraneous load. While this might occur in any complex teaching situation, it might be more likely in procedural settings given the dual-task demands of cognitive and psychomotor learning, and is a potential major hindrance to procedural learning.

Intrinsic versus extraneous load

Most research treats intrinsic and extraneous load as being easy to differentiate from one another. As described earlier, intrinsic load occurs when learners use working memory to accomplish essential task elements, whereas extraneous load occurs when learners use working memory to attend to internal or external stimuli that are not essential to task completion. Put another way, intrinsic load represents productive processing, whereas extraneous load is unproductive processing[18]. However, as Kalyuga and Singh describe, whether a task is essential or non-essential, and whether cognitive load is productive or unproductive, depends on the learning goals[18]. Since learning goals change over time as learners develop competence, certain aspects of procedural learning environments may have differential contributions to cognitive load depending on the learner's level of training. For example, the chaos of the intensive care unit may be considered a distraction to an early medical intern who is only beginning to learn central venous catheter placement technique. However, a critical care fellow who is nearing graduation should be able to manage the multiple stakeholders in the intensive care unit setting, while at the same time placing a central venous catheter. Hence, the procedural environment could be considered a source of extraneous load for the early learner but a source of intrinsic load for the advanced learner, and instructional techniques for the two should differ. In our research we were able to fairly easily parse out intrinsic versus extraneous load, possibly because we mostly focused on early procedural learners. Likewise, few studies reviewed in Chapter 4 focused on the cognitive load of more advanced learners. Because balancing cognitive load and promoting learning among advanced learners is equally important as among early learners, we propose this as a fruitful area for future study.

Germane load

The few studies reviewed in Chapter 4 that focused on germane load were quite limited. Other than the research reported in Chapters 2 and 3, none of the studies investigated factors associated with germane load; the few identified were focused on teaching strategies and design of practice.

Of all the items in the CLIC, we found the germane items most difficult to develop (Chapter 2). Ultimately, we selected wording that reflected perceived mental effort applied to learning the essential demands of colonoscopy. Despite the similarity in wording to the intrinsic load items and positive correlation with intrinsic load, the germane load items loaded separately from intrinsic load, and we found unique factors associated with each (Chapter 3). The lack of association between prior experience or year in training (Chapter 3) suggest that learners of all levels can continue to build and refine procedural skills schemas throughout training, and despite distractions and fatigue (which were not associated with germane load). Chapter 3 also confirms the importance of the teacher in promoting germane load and learning, as supervisor engagement was positively correlated with germane load. Chapter 5 highlights systems issues that may interfere with germane load: inadequate teaching skills and variation in teachers' teaching approaches and demands, while Chapter 6 suggests that excessive verbal teaching can interfere with germane load.

In Chapter 5 teachers reported characteristics of procedural learners that they felt might interfere with germane processing. Some learners were reported to have a goal-orientation that prioritized performance over learning. That is, learners were less concerned with having good technique or understanding how to do a colonoscopy; rather, they just wanted to get the colonoscopy completed quickly and efficiently. Other learners were reported to have low self-efficacy, which reduced their receptiveness to teaching and negatively impacted learning. However, we were not able to investigate how either of these actually affected learner cognitive load (Chapter 6), as we did not consider self-efficacy observable, and we observed no instances of promoting a mastery goal orientation.

Teaching and curricular approaches to promote balanced cognitive load in procedural skills training

This thesis seeks to provide practical recommendations to promote the goals of CLT in procedural settings: to match/optimize intrinsic load to the level of the learner, to minimize extraneous load, and to promote germane load which leads to learning. Unlike many cognitive tasks (such as clinical reasoning), procedural skills usually occur within a discrete period that has a clearly defined beginning and end. We use this as a framework to broadly consider how to promote balanced cognitive load during actual procedural skills teaching sessions (Figure 1). We additionally refer the reader to Table 4 in Chapter 3, and Table 4 in Chapter 3, both of which outline practical recommendations in more granular detail. In this section we will discuss practical recommendations for direct teaching, followed by suggestions for curricular development.

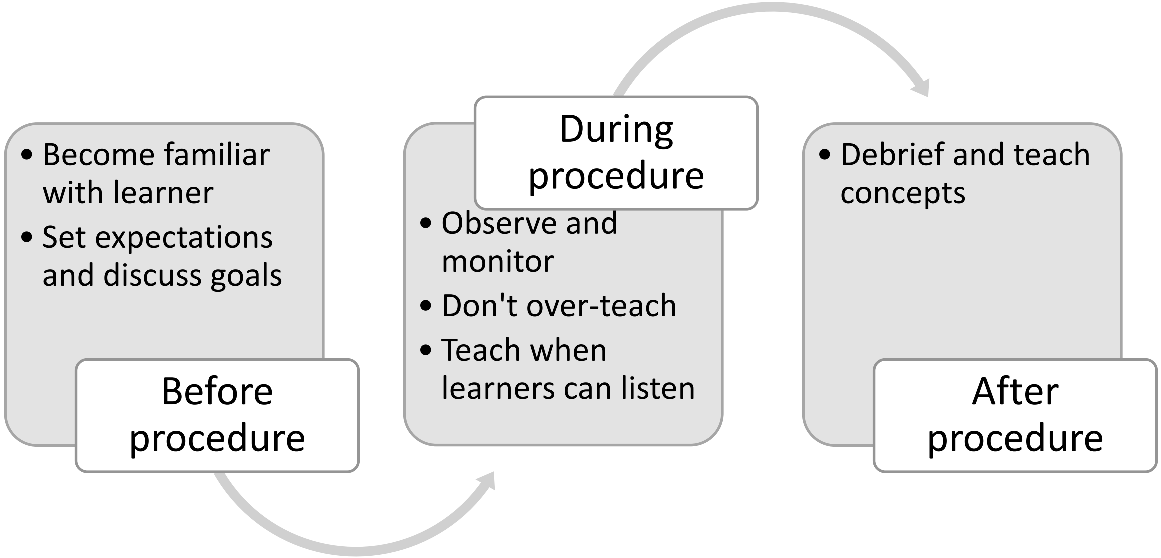


Figure 1. General recommendations to optimize cognitive load before, during and after procedural teaching.

Direct teaching to promote balanced cognitive load

<u>Before procedure: Become familiar with learners' experience and competence</u>

Because procedural skills take a long time to learn and learners progress at different rates, procedural teaching should be modulated based on the learner's experience and competence. In Chapter 6, learners reported that when teachers taught material that was too complex, it contributed to feeling overloaded (likely due to a combination of intrinsic and extraneous load). Yet, when teaching was too simple, it failed to promote germane load. The latter finding likely relates to the expertise reversal effect[56]; that is, "over-teaching" more advanced learners can actually interfere with learning. These learners should be afforded more autonomy (as supported by Chapter 6).

Teachers may have limited familiarity with learner experience and competence in systems where the same teachers and learners do not routinely perform procedures together for extended periods of time. Chapters 5 and 6 underscore methods for faculty to familiarize themselves with their procedural learners' level of expertise, including checking in with the learner prior to a procedural session to ask about experience and procedural progress to-date, discussing learners' progress with colleagues (Clinical Competency Committees as mandated by the Accreditation Council for Graduate Medical Education[61] are a great forum in addition to informal discussion with colleagues), and using routinely collected formative and summative data (e.g., evaluation and examination results). Probably the most efficient among these is a simple check-in before a procedure or procedural session, and a period of observation without intervening (also described in Chapter 5) to assess a learner's procedural skill level. These simple activities may help the teacher understand the learner's abilities, which in turn may guide their teaching along a developmental continuum (Chapter 5), including how they describe concepts, what terminology they use, what portions of the procedure they let the learner do, and how long they permit the learner to struggle. This should benefit intrinsic and germane load goals.

Before procedure: Set expectations and discuss goals

Unmet expectations are a major source of distress in life, and procedural skills training is no exception. When a learner expects to play a major part in a procedure but is relegated to a minor role, it leads to negative emotions that can induce extraneous load and reduce capacity for germane load and learning. Procedural teachers are advised to briefly discuss expectations for each procedure or procedural session, including what portion of the procedure a learner will likely perform, in what circumstances the teacher will likely take over, any portions of the procedure expected to be particularly challenging, and any relevant time constraints[62,63]. Teachers in Chapter 5 reported using this strategy, and learners in Chapter 6 reported it decreased extraneous load and increased germane load. Discussing a learner's goals for a procedure or session can help the teacher know where to target at least some of their teaching; learners in Chapter 6 reported this promoted germane load. The expectation-setting discussion can also illuminate and potentially mitigate competing demands for the learner.

During procedure: Observe and monitor

The associations between teacher engagement and extraneous and germane load (Chapter 3) may owe to teachers observing learners' competence and monitoring learners' degree of cognitive load or overload to modulate their teaching. Teachers in Chapter 5 reported that an initial period of quiet observation to assess a learner's skill level was useful to help guide their teaching along a developmental continuum. Some were able to notice when the learner appeared overwhelmed and used that as a stimulus to adjust their teaching, or to take over the procedure. Learners in Chapter 6 perceived cognitive load benefits when teachers taught to their current stage in learning and when teachers responded to their feelings of cognitive overload. When a procedural teacher is unaware of these factors, teaching is likely to overwhelm the learner or to provide no extrinsic motivation for learning (both described in Chapter 6). The period of observation may also promote a sense of autonomy, which learners felt promoted germane load (Chapter 6).

During procedure: Don't over-teach

Teaching excessively was universally perceived as negatively impacting both extraneous and germane load. Excessive teaching overloads learners' working memories and does not afford them space to develop and test their own solutions, which should promote germane load. Conversely, provision of autonomy was the teaching activity most universally appreciated as promoting learning by learners in Chapter 6, as it provided time and space for them to determine on their own what was happening and to experiment with different methods (both of which promoted germane load) and reduced distractions from the teacher talking excessively (which reduced extraneous load).

<u>During procedure: Teach when learners can listen</u>

As underscored throughout this thesis, procedural training settings place learners at substantial risk of cognitive overload, and so procedural teachers are advised to take advantage of moments when the learner's working memory is not overloaded. Our research identifies two such intra-procedural periods. First is during takeover. Teachers reported using the takeover period as an opportunity to teach (Chapter 5) and learners reported being able to listen and engage in activities promoting germane load during the takeover period (Chapter 6). Setting expectations as described above may diminish negative emotions that learners can experience when the teacher takes over (Chapter 6). Teachers in Chapter 5 also reported advising learners to stop all motor activity and listen to instruction as a way to reduce working memory load and teach, but the impact on learners is unknown, as this strategy was not witnessed during actual observed colonoscopies (Chapter 6). This approach is aligned with the concept of "slowing down when you should" as articulated by Moulton et al[64]. Prompting learners to slow down (or stop) and listen may help cue them in and develop this aspect of expertise.

<u>After procedure: Debrief and teach concepts</u>

The post-procedure period represents a key opportunity to check in with the learner, when they have available working memory capacity to discuss key learning points from the procedure. This is an ideal time to promote beneficial schema formation and refinement, yet seemed underutilized in Chapter 5 (where few participants discussed it) and Chapter 6 (where it was rarely observed). When it did occur in Chapter 6, learners found it universally beneficial for promoting germane load. The post-procedural period is a time where teachers and learners can debrief and discuss multiple aspects of the procedure, including cognitive, psychomotor and clinical decision-making. Spending just a few minutes of time afterwards (as observed in Chapter 6) may contribute substantially to learners' schema formation and refinement. Of course, systems barriers, such as keeping up with a busy clinical schedule, may reduce time available for such debriefing. Additionally, the teacher should assess the learner's readiness to debrief, since learners may experience a period of working memory resource depletion following an intensive learning encouter[50,51].

Curricular design to promote balanced cognitive load

While individual teachers have opportunities to improve cognitive load when teaching procedures, educational leaders may also leverage training curricula, particularly in post-graduate settings, to promote balanced cognitive loads among their learners. Most of these recommendations derive from the studies reviewed in Chapter 4.

Procedural training programs may take several steps to reduce risk of overwhelming intrinsic load, particularly among early procedural learners, as lower level of training was consistently associated with higher intrinsic load. Technical skills training should be scaffolded, starting with partial tasks of low complexity and progressing gradually to partial tasks of higher complexity, and eventually to full authentic tasks. The most clearly articulated curricular design approach that focuses on managing intrinsic load is van Merriënboer's four-component instructional design (4C/ID) method[65]. This method seeks to deconstruct complex tasks to provide a scaffolded approach with progressively more complex tasks and decreasing levels of instructional support, such that the learner can eventually complete an entire complex task autonomously. The four components are: learning tasks (which are sequenced to progress from simple to complex and involve assessments of learning), supportive information (the baseline information that is provided as foundational for all learners, timed developmentally), procedural or "just-in-time" information (which guides learners during actual procedural learning, including corrective feedback and decreases over time), and part-task practice (which promotes deliberate practice leading to automaticity). Several examples of using 4C/ID in cognitive HPE learning settings have been published, including case-presentation skills[66], evidence-based medicine training[67], clinical reasoning in dentistry[68,69], communication skills in nursing[70], and qualitative HPE research methods[71]. We identified only one example of procedural skills training using 4C/ID (related to simulated nephrostomy tube placement training[72]), yet we appreciate great potential impact of studying how 4C/ID could apply to procedural training. Notably, several teaching strategies described by experienced endoscopy teachers (Chapter 5) were well-aligned with 4C/ID. Teaching along a developmental continuum aligns with supportive and procedural information, motor instruction is an example of procedural information, technical assistance and take-over are partial task approaches, and schema teaching provides procedural and supportive information. These are teaching approaches rather than curricular, but we mention them here as they relate to 4C/ID. Additional curricular approaches supported by Chapter 4 to reduce intrinsic load include also relate to the four components of 4C/ID: use of simulation, standardizing common tasks, and scaffolding curricular elements. We recommend further application of 4C/ID to procedural skills curriculum development and assessment.

Curricular approaches to reducing extraneous load largely related to monitoring for, and responding, to learner fatigue, and reducing environmental distractions. Chapter 3, plus two studies in Chapter 4[21,22], documented that fatigue was associated with increased extraneous load, suggesting we should closely monitor procedures performed by learners who are post-call, or who have been working excessive hours, or have other reasons for fatigue. The positive association between queue order and extraneous load (Chapter 3) suggests this is particularly critical near the end of a procedural session. While we found relatively little role for environmental distractions in our empirical research (Chapters 3, 5 and 6), some studies reviewed in Chapter 4 support these as important sources of extraneous load in workplace environments[27,73]. These were not performed in procedural settings, but it is intuitive that such distractions exist in some procedural environments and should be systematically addressed. In one of the few translational studies we identified, reengineering a radiology workplace environment reduced extraneous load and increased satisfaction[74].

Finally, neither our studies nor those reviewed in Chapter 4 addressed curricular or systems approaches to promoting germane load (other than optimizing intrinsic load and reducing extraneous load, which provides space in working memory for activities promoting germane load). However, we speculate that providing learners with supportive information about CLT and metacognition could be useful to promote germane load.

FUTURE DIRECTIONS

The work in this thesis reveals several areas that are important if future research is to further our understanding of how to best teach HPE procedures.

Treatment of CLT in HPE research

We perceive that CLT has been applied to HPE research in a very reductionist manner, based both on the studies reviewed in Chapter 4, and other CLT research in non-workplace HPE settings. A great deal of research has focused on quantifying cognitive load and its subtypes and on simple characterization of the overall cognitive load associated with two or more learning conditions. Conversely, little research has investigated how we should actually be thinking about CLT within HPE, particularly in procedural skills and other workplace settings. Little research has investigated what factors impact cognitive load, how learners experience cognitive load, whether teachers understand cognitive load, and how we can leverage systems, curricula and direct teaching to optimize cognitive load. We perceive that other theoretical frameworks are treated much more broadly in HPE research. The reductionist treatment of CLT may owe to its underpinnings in cognitive psychology, which tends to be experimentally-oriented (indeed, nearly 50% of studies in Chapter 4 were experimental!), and the inherently quantifiable nature of CLT.

While we agree that measuring cognitive load is very important (hence Chapter 2), we appreciate a need to progress to more qualitative and mixed methods research approaches to better understand what CLT means and looks like in complex HPE workplace settings in a more constructivist way. Little such research exists – only 10% of studies in Chapter 4 used mixed methods or qualitative designs. This was the impetus for Chapters 5 and 6, which are unlike any study we identified in Chapter 4, and, we believe, make significant progress to understanding CLT in procedural skills settings. We additionally propose more translational studies to assess impact of curricular innovations and teaching strategies on learners' cognitive load.

Furthermore, CLT provides a very specific and relatively narrow lens through which to view HPE learning, and it could be informed by considering relationship between CLT and other learning theories. In particular, as we learn more about how emotion and relationships impact cognitive load[51,75-77], sociocultural learning theory[78] become relevant. We appreciate potential benefit of approaching "wicked problems" in medical education[79] by simultaneously considering perspectives of multiple learning theories[80].

Additionally, while the research in this thesis focuses on how procedures are taught and learned from a cognitive perspective, a burgeoning body of literature addresses how cognitive processes are embodied within actual physical learning environments. Theories of embodied cognition emphasize that, because of the interconnectedness of the mind, body and environment, bodily experiences influence cognitive processes[81]. In other words, the mind does not process cognitive information in a vacuum, but, rather, within the context of sensorimotor experiences which occur in physical learning environments[82]. This suggests that learners' struggles to understand complex psychomotor maneuvers, or how to interpret and act upon tactile feedback from a surgical instrument or endoscope, are not purely cognitive challenges, but are challenges of embodied cognition. Procedural environments offer among the most intense sensorimotor experiences within all of HPE, and those physical experiences likely affect cognitive learning. Our interviews with endoscopy teachers (Chapter 5) and learners (Chapter 6) suggest this to be the case. We agree with van der Schaaf, Bakker and ten Cate that procedural learners need to develop and refine "motor-action neural circuits"[81], and we further propose that development of such circuits probably represents a form of germane load. Overall, we appreciate strong potential synergy between CLT and embodied cognition, and we hope future research will integrate these two theoretical frameworks.

Identifying overloaded learners

Through the course of the research in this thesis, we have discussed CLT with numerous teachers and learners in various HPE settings, and we have found that most seem to resonate with its central concepts and tenets. We recently designed and implemented a workshop in which teachers applied CLT to their own HPE teaching settings and the variety of situations in which they perceived relevance was extraordinary (data currently being analyzed). In other words, we have found that CLT makes sense to HPE teachers and learners, and yet we do know very little about how to practically identify overloaded learners during actual procedural training. In Chapter 5, teachers described a few cues that learners were overloaded – repeating the same motor movement without success, not trying anything new, lack of responsiveness to instruction, verbal utterances, and purposeless motor movements – yet none of them articulated specifically seeking to identify overloaded learners. In Chapter 6, no learners reported that they felt overloaded, though the observing author (the author of this thesis) anecdotally perceived evidence of overload during some observations (including body language and verbal utterances). In Chapter 4, only one study addressed learner awareness of their own cognitive overload[83].

Learners and teachers alike must be able to monitor for, recognize, articulate, and act upon cognitive overload during procedural skills training. This is crucial to meet needs of all stakeholders – not only learners, but also teachers, patients, and healthcare systems. Research has suggested that secondary task and physiologic approaches may be used to identify overloaded learners[84-86], but such approaches are generally not practical for actual procedural skills training in the workplace.

Hence, we recommend that future research focus on: how teachers and learners can recognize when learners are cognitively overloaded, ways for teachers and learners to communicate overload, and teaching approaches (beyond simply taking over, though that is sometimes needed) to reduce and rebalance cognitive load. We suspect that a qualitative approach is the best way to initiate this line of research. Additionally, though some modern HPE curricula are beginning to teach HPE students about the learning sciences[87], most HPE learners and teachers likely have little to no training regarding how they should learn. Teaching learners and teachers about CLT and metacognition[88], and how to apply these concepts to HPE teaching and learning (both procedural and otherwise), could help both teachers and learners identify when cognitive overload occurs.

Learning and performance

Scholars propose that studying CLT without measures of learning and performance diminishes theoretical and practical impact[43,44,46]. While this viewpoint has legitimacy, we suggest it may derive from the cognitive psychology origins of CLT. In psychology research, experimental designs are of paramount importance, study subjects (i.e., undergraduate students) are plentiful, and tasks for which participants will have little prior experience are readily available. None of these conditions is easily met in HPE settings, particularly procedural settings. Experimental studies necessarily require the ability to control most aspects of the study environment. In HPE, such control is possible only in simulated settings; workplace settings have numerous stakeholders that prevent altering the setting to meet experimental demands. Although there is benefit from studying simulation, we assert that most procedural skills simulations are unlikely to closely approximate the actual cognitive and sociocultural complexity of a real procedural setting; this assertion is supported by the very narrow scope and objectives of most of the studies reviewed in Chapter 4, 78% of which were entirely simulation-based. Furthermore, with each progressive level of training, the number of trainees in a given program diminishes. Consider for example, the training program with which the author of this thesis is affiliated. The University of California San Francisco has one of the largest gastroenterology training programs in the United States, which amounts to only 6 trainees per class. It is for this reason that we sought a national subject pool for Chapters 2 and 3, yet such large-scale studies are largely limited to survey-based designs. Multi-center efforts can increase participant numbers, but such collaborations are challenging to achieve. Finally, because HPE learners rapidly progress in knowledge and experience, it is difficult to recruit participants who are truly novices for whom the learning or performance impact of experimental conditions can be reliably attributed, but who are ethically permitted to perform a real procedure on an actual patient.

Considering these challenges of HPE research, we return to the question of learning and performance. We assert that the body of work detailed in this thesis, including the review in Chapter 4 (which includes 116 studies), plus other research not performed or described here, clearly supports the relevance of CLT to HPE teaching, including procedural skills training. We further assert that the theoretical basis for CLT and the primary identities of intrinsic, extraneous and germane load (or working memory resource allocated to dealing with intrinsic load[43]) are well-supported as defined in this thesis and by other research[23,38,48,54,89-91]. Therefore, although measures of performance and learning are certainly of interest, we argue that they are not mandatory for workplace HPE research in studies that are otherwise well-designed.

For those interested in learning and performance as they relate to CLT, one way to study them in authentic procedural training settings would be to assess cognitive load (ideally subtypes) and performance longitudinally over the course of training to determine how levels of intrinsic, germane and extraneous load are associated with learning and performance. However, such a study design would have to be quite focused and reductionist, and if no specific teaching strategies are studied, the practical implications might be limited.

STRENGTHS AND LIMITATIONS

The research in this thesis has a number of strengths. It follows a logical progression from measurement of cognitive load to factors associated with cognitive load to teachers' idealized strategies to actual teaching activity use and learner impact; in other words, from theoretical to practical. The BEME review is intentionally positioned as the middle study to help translate theoretical knowledge to more applied, practical studies. The research addresses CLT and procedural skills training from multiple viewpoints, including learners (Chapters 2, 3 and 6), teachers (Chapter 5) and the literature (Chapter 4) and using diverse methods (quantitative, qualitative, and mixed methods). Chapters 2 and 3 involve what can truly be considered a nationally representative sample of gastroenterology trainees in the United States, while Chapters 5 and 6 hone in on in-depth experiences of procedural teachers and learners. The in-depth observations in Chapter 6 provide a unique vantage point to link actual teaching to perceived cognitive load impact. Finally, we strove for a high level of theory integration[92] throughout, supporting validity of both theoretical and practical implications that we derived.

There are limitations to this research as well. Other than Chapter 4, we focused on a single procedural setting – gastrointestinal endoscopy. However, we considered the study of numerous different procedural settings to be too diffuse and heterogeneous to yield firm results. As we have argued, gastrointestinal endoscopy is an ideal exemplar procedural setting in which to study CLT, for numerous reasons, and we believe that most lessons learned are readily generalizable to a wide variety of procedural settings. Additionally, Chapter 4 delivers contributions from a wide range of training settings, both procedural and cognitive, within and outside of the health professions. We measured cognitive load subtypes using a post-hoc summative approach (Chapters 2, 3 and 6). This was imperative to accomplish our research goals given the great importance of understanding the cognitive load subtypes, and it was the only method feasible for recruiting a large nationally representative sample of learners. Furthermore, since all cognitive load measurement methods have limitations, and there is no gold standard, as discussed earlier, we find this approach to reasonable. We did not collect measures of performance or learning, yet as described above we do not feel these to be crucial. Finally, we focused on procedural skills, yet from a cognitive perspective, and so we did not closely examine how actual technical skills are developed, refined, used and taught. There is a vast literature addressing technical skill development[93-102], much of which focuses on very specific specialties and tasks. We believe our approach using CLT to inform procedural skills training provides broader implications more readily applicable to a wide variety of procedural training settings.

CONCLUSION

This thesis aims to illuminate how cognitive load theory can be used to understand the complexities of teaching and learning procedural skills in the health professions. It argues that procedural training is a unique and challenging undertaking within health professions education that has high potential for cognitive overload in learners. The work detailed here addresses multiple issues that were previously unknown including how to measure intrinsic, extraneous and germane load during an exemplar procedural training setting, what factors are associated with each cognitive load subtype during procedural training, how experienced endoscopy teachers' teaching strategies align with cognitive load theory, how endoscopy trainees experience actual colonoscopy teaching, and how that teaching affects their perceived cognitive load. Additionally, the review presented in Chapter 4 assimilates and synthesizes a vast amount of literature to distill multiple recommendations for direct teaching, curricular design, learning environment and metacognition relevant to procedural skills training. We believe that cognitive load theory has the potential to further inform and improve procedural skills training, and we hope that its use can reduce the distress that arises from cognitive overload as described in the Preface.

REFERENCES

1. Lipsman N, Khan O, Kulkarni AV. “The Actualized Neurosurgeon”: A Proposed Model of Surgical Resident Development. *World Neurosurg.* 2017;99:381-386.
2. Esterl RM, Henzi DL, Cohn SM. Senior Medical Student “Boot Camp”: Can Result in Increased Self-Confidence Before Starting Surgery Internships. *Current Surgery.* 2006;63(4):264-268.
3. Cassel CK, Reuben DB. Specialization, subspecialization, and subsubspecialization in internal medicine. *N Engl J Med.* 2011;364(12):1169-1173.
4. Sorensen MJ. Surgical subspecialization: escape route for surgeons or added benefit for patients? *J Grad Med Educ.* 2014;6(2):215-217.
5. van Zanten M, Norcini JJ, Boulet JR, Simon F. Overview of accreditation of undergraduate medical education programmes worldwide. *Med Educ.* 2008;42(9):930-937.
6. Linville MD, Jr., Bates JE. Graduate Medical Education-Accelerated Change. *Am J Med Sci.* 2017;353(2):126-131.
7. Weinstein DF. The Elusive Goal of Accountability in Graduate Medical Education. *Acad Med.* 2015;90(9):1188-1190.
8. Attrill S, Lincoln M, McAllister S. Supervising international students in clinical placements: perceptions of experiences and factors influencing competency development. *BMC Med Educ.* 2016;16:180.
9. Atkinson RCS, R.M. Human memory: A proposed system and its control processes. In: Spence KWTS, J., ed. *Psychology of Learning and Motivation.* New York: Academic Press; 1968:89-195.
10. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci.* 1988;12:257-285.
11. Young JQ, van Merriënboer J, Durning S, ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Med Teach.* 2014;36(5):371-384.
12. Naismith LM, Cavalcanti RB. Validity of Cognitive Load Measures in Simulation-Based Training: A Systematic Review. *Acad Med.* 2015;90(11 Suppl):S24-35.
13. Young JQ, Sewell JL. Applying cognitive load theory to medical education: construct and measurement challenges. *Perspect Med Educ.* 2015;4(3):107-109.
14. Lau KH, Moeller J. Reliable cognitive load measurement using psychometrics: towards a model of objective teaching programme optimisation. *Med Educ.* 2016;50(6):606-609.
15. DeLeeuw KE, Mayer RE. A comparison of three measures of cognitive load: evidence for separable measures of intrinsic,

extraneous, and germane load. *J Educ Psych.* 2008;100(1):223-234.
16. Eysink THS, de Jong T, Berthold K, Kolloffel B, Opfermann M, Wouters P. Learner Performance in Multimedia Learning Arrangements: An Analysis Across Instructional Approaches. *Am Educ Res J.* 2009;46(4):1107-1149.
17. Paas F, Tuovinen JE, Tabbers H, van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. *Educ Psychol.* 2003;38(1):63-71.
18. Kalyuga S, Singh AM. Rethinking the Boundaries of Cognitive Load Theory in Complex Learning. *Educ Psychol Rev.* 2016;28(4):831-852.
19. Hart SG. NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proc Hum Factors Ergon Soc Annu Meet.* 2006;50(9):904-908.
20. Artino AR, Jr., La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach.* 2014;36(6):463-474.
21. Bertram DA, Hershey CO, Opila DA, Quirin O. A measure of physician mental work load in internal medicine ambulatory care clinics. *Med Care.* 1990;28(5):458-467.
22. Bertram DA, Opila DA, Brown JL, et al. Measuring physician mental workload: reliability and validity assessment of a brief instrument. *Med Care.* 1992;30(2):95-104.
23. Dankbaar ME, Alsma J, Jansen EE, van Merriënboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ Theory Pract.* 2016;21(3):505-521.
24. Gaba DM, Lee T. Measuring the workload of the anesthesiologist. *Anesth Analg.* 1990;71(4):354-361.
25. Mulcock PM, Grassley J, Davis M, White K. Beyond the Dedicated Education Unit: Using Cognitive Load Theory to Guide Clinical Placement. *J Nurs Educ.* 2017;56(2):105-109.
26. Naismith LM, Cheung JJH, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ.* 2015;48(8):805-814.
27. Tremblay ML, Lafleur A, Leppink J, Dolmans DH. The simulated clinical environment: Cognitive and emotional impact among undergraduates. *Med Teach.* 2017;39(2):181-187.
28. Weinger MB, Vredenburgh AG, Schumann CM, et al. Quantitative description of the workload associated with airway management procedures. *J Clin Anes.* 2000;12(4):273-282.
29. Young JQ, Boscardin CK, van Dijk SM, et al. Performance of a cognitive load inventory during simulated handoffs: Evidence for validity. *SAGE Open Med.* 2016;4:2050312116682254.
30. Young JQ, Irby DM, Barilla-LaBarca ML, ten Cate O, O'Sullivan PS. Measuring cognitive load: mixed results from a handover simulation

for medical students. *Perspect Med Educ.* 2016;5(1):24-32.
31. Dahlstrom N, Nahlinder S. Mental workload in aircraft and simulator during basic civil aviation training. *Int J Aviat Psyc.* 2009;19(4):309-325.
32. Murai K, Ishikura A, Hayashi Y, et al. Toward evaluation of ship bridge teamwork for marine simulator training based on heart rate variability. *World Automation Congress.* 2010:1-6.
33. Leppink J, Paas F, van der Vleuten CP, van Gog T, van Merriënboer JJ. Development of an instrument for measuring different types of cognitive load. *Behav Res Method.* 2013;45(4):1058-1072.
34. Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RS. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World J Surg.* 2011;35(9):1961-1969.
35. Durning SJ, Costanzo M, Artino AR, Jr., et al. Functional Neuroimaging Correlates of Burnout among Internal Medicine Residents and Faculty Members. *Front Psychiatry.* 2013;4:131.
36. Bahrami P, Schweizer TA, Tam F, Grantcharov TP, Cusimano MD, Graham SJ. Functional MRI-compatible laparoscopic surgery training simulator. *Magn Reson Med.* 2011;65(3):873-881.
37. Shachak A, Hadas-Dayagi M, Ziv A, Reis S. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. *J Gen Intern Med.* 2009;24(3):341-348.
38. Klepsch M, Schmitz F, Seufert T. Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Front Psychol.* 2017;8:1997.
39. Naismith LM, Haji FA, Sibbald M, Cheung JJ, Tavares W, Cavalcanti RB. Practising what we preach: using cognitive load theory for workshop design and evaluation. *Perspect Med Educ.* 2015;4(6):344-348.
40. Plass JL, Moreno R, Brunken R. Cognitive Load Theory. In: *Cognitive Load Theory.* New York, NY: Cambridge University Press; 2010:17.
41. Sweller J, van Merriënboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. *Educ Psychol Rev.* 1998;10(3):251-296.
42. Kalyuga S. Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educ Psychol Rev.* 2011;23(1):1-19.
43. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. *Perspect Med Educ.* 2015;4(3):119-127.
44. Leppink J. Cognitive load theory: Practical implications and an important challenge. *J T U med sc.* 2017;12(5):385-391.
45. Leppink J. Cognitive load measures mainly have meaning when they are combined with learning

outcome measures. *Med Educ.* 2016;50(9):979.
46. Naismith LM, Cavalcanti RB. Measuring germane load requires correlation with learning. *Med Educ.* 2017;51(2):228.
47. Leppink J, Paas F, van Gog T, van der Vleuten CPM, van Merriënboer JJ. Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learn Instruct.* 2014;30:32-42.
48. Debue N, van de Leemput C. What does germane load mean? An empirical contribution to the cognitive load theory. *Front Psychol.* 2014;5:1099.
49. Kelley P. Spaced learning: a harbringer of a paradigm shift in education. *Education Today.* 2009;59:3-6.
50. Chen O, Castro-Alonso JC, Paas F, Sweller J. Extending Cognitive Load Theory to Incorporate Working Memory Resource Depletion: Evidence from the Spacing Effect. *Educ Psychol Rev.* 2018;30(2):483-501.
51. Sweller J, van Merriënboer JJG, Paas F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ Psychol Rev.* 2019; Jan 22:1-32 [Epub ahead of print].
52. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830-837.
53. Kane MT. Current concerns in validity theory. *J Educ Meas.* 2001;38(4):319-342.
54. Chen R, Grierson L, Norman G. Manipulation of cognitive load variables and impact on auscultation test performance. *Adv Health Sci Educ Theory Pract.* 2015;20(4):935-952.
55. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010;44(1):85-93.
56. van Merriënboer JJ, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev.* 2005;17(2):147-177.
57. Kaltenbach T, Leung C, Wu K, Yan K, Friedland S, Soetikno R. Use of the colonoscope training model with the colonoscope 3D imaging probe improved trainee colonoscopy performance: a pilot study. *Dig Dis Sci.* 2011;56(5):1496-1502.
58. Mansukhani MP, Kolla BP, Surani S, Varon J, Ramar K. Sleep deprivation in resident physicians, work hour limitations, and related outcomes: a systematic review of the literature. *Postgrad Med.* 2012;124(4):241-249.
59. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA.* 2011;306(9):978-988.
60. Ayres P, Paas F. Making instructional animations more effective: a cognitive load

approach. *Appl Cogn Psychol.* 2007;21(6):695-700.
61. Hauer KE, Chesluk B, Iobst W, et al. Reviewing residents' competence: a qualitative study of the role of clinical competency committees in performance assessment. *Acad Med.* 2015;90(8):1084-1092.
62. Dilly CK, Sewell JL. How to Give Feedback During Endoscopy Training. *Gastroenterology.* 2017;153(3):632-636.
63. Waschke KA, Anderson J, Macintosh D, Valori RM. Training the gastrointestinal endoscopy trainer. *Best Pract Res Clin Gastroenterol.* 2016;30(3):409-419.
64. Moulton CA, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med.* 2007;82(10 Suppl):S109-116.
65. Vandewaetere M, Manhaeve D, Aertgeerts B, Clarebout G, van Merriënboer JJ, Roex A. 4C/ID in medical education: How to design an educational program based on whole-task learning: AMEE Guide No. 93. *Med Teach.* 2015;37(1):4-20.
66. Daniel M, Stojan J, Wolff M, et al. Applying four-component instructional design to develop a case presentation curriculum. *Perspect Med Educ.* 2018;7(4):276-280.
67. Maggio LA, Cate OT, Irby DM, O'Brien BC. Designing evidence-based medicine training to optimize the transfer of skills from the classroom to clinical practice: applying the four component instructional design model. *Acad Med.* 2015;90(11):1457-1461.
68. Postma TC, White JG. Developing clinical reasoning in the classroom - analysis of the 4C/ID-model. *Eur J Dent Educ.* 2015;19(2):74-80.
69. Postma TC, White JG. Developing integrated clinical reasoning competencies in dental students using scaffolded case-based learning - empirical evidence. *Eur J Dent Educ.* 2016;20(3):180-188.
70. Susilo AP, van Merriënboer J, van Dalen J, Claramita M, Scherpbier A. From lecture to learning tasks: use of the 4C/ID model in a communication skills course in a continuing professional education context. *J Contin Educ Nurs.* 2013;44(6):278-284.
71. Holtslander LF, Racine L, Furniss S, Burles M, Turner H. Developing and piloting an online graduate nursing course focused on experiential learning of qualitative research methods. *J Nurs Educ.* 2012;51(6):345-348.
72. Tjiam IM, Schout BM, Hendrikx AJ, Scherpbier AJ, Witjes JA, van Merriënboer JJ. Designing simulator-based training: an approach integrating cognitive task analysis and four-component instructional design. *Med Teach.* 2012;34(10):e698-707.
73. Durning SJ, Artino AR, Boulet JR, Dorrance K, van der Vleuten C, Schuwirth L. 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.* 2012;17(1):65-79.
74. Lee MH, Schemmel AJ, Pooler BD, et al. Radiology Workflow Dynamics: How Workflow Patterns Impact Radiologist Perceptions of Workplace Satisfaction. *Acad Radiol.* 2017;24(4):483-487.
75. Fraser K, Huffman J, Ma I, et al. The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. *Chest.* 2014;145(5):958-963.
76. Fraser K, Ma I, Teteris E, Baxter H, Wright B, McLaughlin K. Emotion, cognitive load and learning outcomes during simulation training. *Med Educ.* 2012;46(11):1055-1062.
77. Fraser K, McLaughlin K. Temporal pattern of emotions and cognitive load during simulation training and debriefing. *Med Teach.* 2018; Apr 24:1-6 [Epub ahead of print].
78. Kahlke R, Bates J, Nimmon L. When I say … sociocultural learning theory. *Medical Education.* 2019;53(2):117-118.
79. Hawick L, Cleland J, Kitto S. Getting off the carousel: Exploring the wicked problem of curriculum reform. *Perspect Med Educ.* 2017;6(5):337-343.
80. Kamel-ElSayed S, Loftus S. Using and Combining Learning Theories in Medical Education. *Med Sci Educ.* 2018;28(1):255-258.
81. van der Schaaf M, Bakker A, ten Cate O. When I say ... embodied cognition. *Med Educ.* 2018; Sep 26 [Epub ahead of print].
82. Galetzka C. The Story So Far: How Embodied Cognition Advances Our Understanding of Meaning-Making. *Front Psychol.* 2017;8:1315.
83. Moos DC, Pitton D. Student Teacher Challenges: Using the Cognitive Load Theory as an Explanatory Lens. *Teaching Education.* 2014;25(2):127-141.
84. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang GZ, Darzi A. Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surg Endosc.* 2015;29(2):405-413.
85. Rasmussen SR, Konge L, Mikkelsen PT, Sorensen MS, Andersen SA. Notes From the Field: Secondary Task Precision for Cognitive Load Estimation During Virtual Reality Surgical Simulation Training. *Eval Health Prof*. 2016;39(1):114-120.
86. Morales JM, Ruiz-Rabelo JF, Diaz-Piedra C, Di Stasi LL. Detecting Mental Workload in Surgical Teams Using a Wearable Single-Channel Electroencephalographic Device. *J Surg Educ.* 2019; Jan [Epub ahead of print].
87. Zijdenbos I, Fick T, ten Cate O. How we offer all medical students training in basic teaching skills. *Med Teach.* 2011;33(1):24-26.

88. Medina MS, Castleberry AN, Persky AM. Strategies for Improving Learner Metacognition in Health Professional Education. *Am J Pharm Educ.* 2017;81(4):78.
89. Korbach A, Brünken R, Park B. Differentiating Different Types of Cognitive Load: a Comparison of Different Measures. *Educ Psychol Rev.* 2018;30(2):503-529.
90. Cook DA, Castillo RM, Gas B, Artino AR, Jr. Measuring achievement goal motivation, mindsets and cognitive load: validation of three instruments' scores. *Med Educ.* 2017;51(10):1061-1074.
91. Chen R, Dore K, Grierson LEM, Hatala R, Norman G. Cognitive load theory: Implications for nursing education and research. *Can J Nurs Res.* 2014;46(2):28-41.
92. Kumasi KD, Charbonneau DH, Walster D. Theory talk in the library science scholarly literature: An exploratory analysis. *Library & Information Science Research.* 2013;35:175-180.
93. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. *Br J Surg.* 2012;99(10):1322-1330.
94. Jelovsek JE, Kow N, Diwadkar GB. Tools for the direct observation and assessment of psychomotor skills in medical trainees: a systematic review. *Med Educ.* 2013;47(7):650-673.
95. van der Meijden OA, Schijven MP. The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review. *Surg Endosc.* 2009;23(6):1180-1190.
96. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. *Br J Surg.* 2004;91(12):1549-1558.
97. van der Wiel SE, Kuttner Magalhaes R, Rocha Goncalves CR, Dinis-Ribeiro M, Bruno MJ, Koch AD. Simulator training in gastrointestinal endoscopy - From basic training to advanced endoscopic procedures. *Best Pract Res Clin Gastroenterol.* 2016;30(3):375-387.
98. Ekkelenkamp VE, Koch AD, de Man RA, Kuipers EJ. Training and competence assessment in GI endoscopy: a systematic review. *Gut.* 2016;65(4):607-615.
99. Mirza S, Athreya S. Review of Simulation Training in Interventional Radiology. *Acad Radiol.* 2018;25(4):529-539.
100. Mills DM, Williams DC, Dobson JV. Simulation training as a mechanism for procedural and resuscitation education for pediatric residents: a systematic review. *Hosp Pediatr.* 2013;3(2):167-176.
101. Savran MM, Sorensen SM, Konge L, Tolsgaard MG, Bjerrum F. Training and Assessment of Hysteroscopic Skills: A Systematic Review. *J Surg Educ.* 2016;73(5):906-918.

102. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and technical skills performance in health care: a systematic review. *Med Educ.* 2015;49(11):1065-1085.

Summary

Procedural skills training is one of the most challenging undertakings for health professions education (HPE) learners. Medical procedures require learners to develop, refine and apply learning schemas for both cognitive and psychomotor concepts, yet most HPE learners are accustomed to learning primarily in the cognitive domain. This presents a dual-task demand which is challenging in and of itself, yet additional challenges are present as well. Most procedural tasks take many repetitions to learn, and so learners who are accustomed to performing well may frequently fail early in training; this can induce negative emotions. Terminology associated with psychomotor components of procedures may be unfamiliar, making understanding the teacher more challenging. Decisions during procedures must be made without the luxury of extended time to think and discuss, yet decisions made have potential for immediate physical harm to patients. Many procedures are performed in urgent or emergent settings which causes additional stress. Taken together, these and other challenges present an incredibly complex learning environment where learners may become easily overwhelmed.

At the same time, procedural teachers also endure significant challenges. They must prioritize patient care to ensure patients receive safe, efficient and high-quality procedures, yet at the same time they must enable the learner to take on a developmentally appropriate portion of the procedure, provide guidance when needed, use terminology and words that the learner understands, and decide when to take over a procedure. The teacher must also balance needs of other stakeholders including the patient, staff, healthcare systems, and themselves. However, many procedural teachers lack significant training in education and may struggle to balance these important competing demands, while at the same time accomplishing the task at hand.

This thesis seeks to address these and other challenges inherent to procedural teaching and learning in the health professions by considering the lens of cognitive load theory (CLT). CLT is a cognitive learning theory oriented around the Atkinson-Shriffin model of human memory and focused on the limitations of human working memory. As opposed to sensory memory and long-term memory, which are considered infinite, the working memory can only manage a few pieces of information at any given time. The purpose of working memory is to attend to relevant aspects of sensory memory and to transfer that information into long-term memory in tidy packages called learning schema. These schemas, which are cognitive representations of complex topics or constructs, can then be retrieved from long-term memory for use in the future. From a cognitive perspective, learning occurs through formation and refinement of cognitive schemas, and more complex and refined schemas are associated with greater expertise and automaticity.

The combined activities of the working memory are referred to as “cognitive load”. The working memory may engage in three general activities, which are the cognitive load

subtypes. *Intrinsic load* occurs when learners accomplish essential task elements. Intrinsic load is greater when tasks are more complex or when learners are less experienced. When learners use working memory resources to package informational elements from intrinsic load into learning schema, *germane load* occurs. Germane load is ultimately what leads to learning and automaticity. *Extraneous load* occurs whenever learners use working memory resources in unproductive ways, such as attending to external or internal distractions, or trying to understand unclear or confusing teaching or instructions. Extraneous load never contributes to learning. When the combined amount of intrinsic, germane and extraneous cognitive load exceeds working memory capacity, cognitive overload develops, negatively impacting learning and performance. According to CLT, learning is optimized when extraneous load is minimized and intrinsic load matched to the learner's prior experience and knowledge; in the motivated learner this promotes germane load, which leads to learning. CLT has been well-studied in classrooms and experimental cognitive psychology settings, yet there is burgeoning interest in applying it to HPE workplace environments.

CLT seems an ideal theoretical lens through which to study procedural skills training, given that the material to be learned is complex (as described above) and there is substantial risk for cognitive overload. To date, however, study of CLT in procedural settings has been limited. Because we appreciated substantial potential benefit of CLT for procedural learning, we designed the series of studies detailed in this thesis. We selected gastrointestinal (GI) endoscopy as an exemplar procedural setting for study given its status as a "middle-of-the-road" procedure in terms of complexity, duration and learnability, and because of its significant impact on the health of the public. We designed studies to yield implications that would be generalizable to a wide variety of procedural settings.

Chapter 1 provides a detailed overview of the challenges of procedural skills training, theoretical tenets of CLT, a rationale for using CLT as a lens for studying procedural skills training, and a rationale for studying GI endoscopy as an exemplar procedural skills setting.

Because cognitive load exists in the working memory, we needed to identify a method to measure or estimate it for our studies. The most common approach is to use a single-item estimate of mental effort, but we appreciated the critical importance of estimating the cognitive load subtypes (intrinsic, extraneous and germane load) in order to develop theoretical and practically relevant conclusions. We therefore started the thesis research by developing a psychometric instrument – the Cognitive Load Inventory for Colonoscopy (CLIC) – to measure intrinsic, extraneous and germane load during colonoscopy training **(Chapter 2)**. We followed a rigorous systematic process to develop the CLIC and then tested it among a nationally representative sample of 477 colonoscopy learners in the United States. Chapter 2 presents extensive evidence for validity of the CLIC. We tested 19 items (8 for intrinsic load, 6 for extraneous load, 5 for germane load) and retained 15 in the final model (7 for intrinsic load, 4 for extraneous load, 4 for germane load). Exploratory factor analysis identified a three-factor solution consistent with intrinsic, extraneous and germane load. All items fit within the expected factor with excellent factor loading. Confirmatory factor analysis yielded satisfactory fit parameters. Chapter 2 concludes that the CLIC measures three constructs among colonoscopy learners, which we believe represent intrinsic, extraneous and germane cognitive load.

We next sought to explore factors associated with each of the cognitive load subtypes. To accomplish this, in **Chapter 3**, we analyzed a large body of additional data obtained from the 477 participants in Chapter 2. We queried participants regarding multiple features of themselves (i.e., the learner), the task/patient, the procedural setting/environment, and the supervisor (teacher). We performed multivariate modeling to identify features associated with intrinsic, extraneous and germane load, developing unique models for each that were theoretically consistent with CLT. Intrinsic load was primarily associated with features of learners (negatively associated with year in training and prior experience, positively associated with fatigue) and patients/tasks (positively associated with poor patient tolerance of procedure and number of maneuvers performed). Extraneous load was primarily associated with features of the setting (positively associated with queue order) and the supervisor (negatively associated with supervisor confidence and engagement, positively associated with supervisor taking over the procedure), and also positively associated with learner fatigue. Germane load was positively associated with supervisor engagement. The different patterns of features associated with each subtype of cognitive load made sense theoretically, and also provided additional evidence for validity of the CLIC. While the research in Chapter 3 was performed among colonoscopy learners, the features studied are readily generalizable to diverse procedural settings.

Having learned how to measure cognitive load during GI endoscopy training, and armed with a preliminary understanding of what might impact cognitive load in procedural settings, we sought to direct our efforts towards how specific teaching approaches might impact cognitive load. To accomplish this, we wanted to leverage the existing body of relevant literature, and so we designed a scoping review of studies that have used CLT concepts to study workplace training of professional learners both within and outside of the health professions **(Chapter 4)**. This was performed as a Best Evidence in Medical Education (BEME) review. Included studies tended to be narrowly focused, and cognitive load subtypes were often not reported, yet the 116 diverse studies yielded multiple theoretical implications and practical recommendations, which Chapter 4 discusses in great detail. Importantly, those studies informed development of the final two thesis studies.

While scholars have proposed and studied teaching strategies that can be used to promote goals of CLT, few studies reviewed in Chapter 4 addressed which teaching strategies could be leveraged to improve cognitive load specifically in HPE procedural training. We wanted to know how experts approached procedural teaching, and so we designed a qualitative study interviewing 16 experienced endoscopy teachers in the United States, Canada and the Netherlands **(Chapter 5)**. We purposively sampled participants with deep insight into endoscopy teaching and educational approaches. We asked them about struggles they saw their endoscopy learners experience and teaching strategies they used to help learners overcome those challenges. We analyzed transcripts using template analysis. Participants described learner challenges which we categorized as relating to learners (performance over mastery goal orientation, low self-efficacy, lack of awareness); the task (psychomotor challenges, mental model development, tactile understanding); the teacher (teacher-trainee relationship, inadequate teaching, variability in teaching); and the setting (internal and external distractions, systems issues). Participants reported using teaching strategies that were aligned with CLT, including those benefitting intrinsic load by matching to the learner level (teaching along developmental continuum, motor instruction, technical assistance), minimizing extraneous load (optimize environment, systems solutions, emotional support, define expectations), and promoting germane load (promote mastery, teach schemas, stop and focus).

Finally, we sought to understand how these and other teaching activities were enacted in actual colonoscopy teaching among 'everyday' endoscopy teachers, and what impact learners perceived the teaching as having on their intrinsic, germane and extraneous cognitive load. To accomplish this, we performed a mixed methods study, observing actual colonoscopies performed by gastroenterology fellows (learners) and attending physicians (teachers) at two large hospitals **(Chapter 6)**. We selected 11 of the teaching activities from Chapter 5 that we expected to be observable: motor instruction, feedback, schema teaching, checking understanding, takeover, setting expectations, technical assistance, emotional support, environment modification, promote mastery goal orientation, stop and listen. We used event sampling and real-time content analysis to determine when those 11 teaching activities occurred and how they were enacted. Afterwards we interviewed each fellow to explore how they perceived each observed activity as impacting cognitive load subtypes and analyzed transcripts using content analysis. We also correlated instances of germane load promoting activities with measured germane load. We observed 515 instances of the 11 teaching activities among 10 colonoscopies performed by 10 different attendings and 8 different fellows. Intensity of teaching varied widely from 0.7 to 3.3 teaching activities per minute procedural time. The pre- and post-procedure time periods were rarely used for teaching. The most commonly observed teaching activities were motor instruction, feedback, schema teaching and checking understanding. Two new teaching activities were characterized: providing cognitive information and affording autonomy. Most teaching activities were perceived to have benefits of reducing intrinsic load and increasing germane load; few impacted extraneous load. However, even beneficial teaching activities were perceived to induce extraneous load when the attending taught excessively. Instances of germane load-promoting activities were moderately positively correlated with measured germane load (Spearman's rho was 0.44). We conclude that teaching activities may be leveraged to promote goals of CLT during procedural skills training, but that they must be used thoughtfully to avoid cognitively overloading learners.

The thesis concludes with the Discussion in **Chapter 7**, which synthesizes the work in this thesis with other literature related to CLT and procedural skills training. Chapter 7 specifically addresses several major contributions of this work to the literature including: measurement of cognitive load subtypes, conceptualization of germane load, what contributes to intrinsic, germane and extraneous load in procedural skills training settings, and how to leverage teaching and curricular design to balance and optimize procedural learners' cognitive load. Directions for future research are discussed, particularly as they relate to treatment of CLT in HPE research, understanding the phenomenon of cognitive overload, and the relevance of measuring learning and performance in HPE studies of CLT. Finally, we address strengths and limitations of the thesis work, and argue for generalizability of findings to diverse procedural settings.

Overall this thesis makes substantial theoretical and practical contributions to our understanding of teaching procedural skills in the health professions through the lens of CLT. We hope that is serves as a stimulus for further research and teaching and curricular innovations as we strive to optimize learning in this challenging yet rewarding domain in the health professions.

Samenvatting

Training in procedurele vaardigheden is een van de meest uitdagende onderdelen van het onderwijs voor de beroepen in de gezondheidszorg (HPE). Procedurele vaardigheidstraining vereist van lerenden, die vooral gewend zijn kennis tot zich te nemen, dat ze cognitieve en psychomotorische leerschema's ontwikkelen, verfijnen en toepassen. Dit is een dubbele taak ('dual task'), die op zichzelf al uitdagend is, maar er zijn ook extra uitdagingen. De meeste procedurele taken vragen vele herhaling om ze te leren, en lerenden die gewend zijn om in andere opzichten goed te presteren kunnen soms vroeg in vaardigheidstraining al falen; met alle negatieve emoties van dien. Termen die gebruikt worden bij vaardigheids-training kunnen onbekend zijn, wat het begrijpen van de docent bemoeiijkt. Beslissingen tijdens live procedures gebeuren snel, zonder de luxe van tijd om na te denken en zaken te bespreken, en kunnen onmiddellijk lichamelijk letsel bij patiënten teweeg brengen. Veel procedures worden uitgevoerd in acute omstandigheden die extra stress veroorzaken. Alles bij elkaar bieden deze en bijkomende uitdagingen een ongelooflijk complexe leeromgeving waarin lerenden gemakkelijk overweldigd kunnen raken.

Tegelijkertijd hebben ook docenten bij live vaardigheidstraining (met live wordt hier bedoeld in de kinische praktijk) grote uitdagingen. Ze moeten voorrang geven aan patiëntenzorg om ervoor te zorgen dat patiënten veilige, efficiënte en hoogwaardige ingrepen ondergaan, maar tegelijkertijd moeten ze de lerende in staat stellen onderdelen van de ingreep op zich te nemen die nodig zijn voor hun ontwikkeling, begeleiding bieden wanneer dat nodig is, terminologie gebruiken die de lerende begrijpt en beslissen wanneer een verrichting moet worden overgenomen. De docent moet ook rekening houden met de de belangen van anderen, zoals de patient en medewerkers, het ziekenhuis en zichzelf. Veel live vaardigheidsdocenten ontberen echter zelf de noodzakelijke training in onderwijs en kunnen moeite hebben om al deze belangrijke concurrerende eisen in evenwicht te brengen, terwijl ze tegelijkertijd hun onderwijstaak vervullen.

Dit proefschrift is gericht op de uitdagingen voor lerenden en docenten in live vaardigheidsonderwijs voor beroepen in de gezondheidszorg vanuit het perspectief van de cognitieve belastingstheorie (Cognitive Load Theory of CLT) te beschouwen. CLT is een cognitieve leertheorie gebaseerd op het Atkinson-Shriffin-model van het menselijk geheugen en gericht op de beperkingen van het menselijk werkgeheugen. In tegenstelling tot sensorisch geheugen en langetermijngeheugen, waarvan de capaciteit als oneindig worden beschouwd, kan het werkgeheugen slechts een paar stukjes informatie op elk willekeurig moment bewerken. Het doel van het werkgeheugen is om relevante aspecten van het sensorisch geheugen te verwerken en die informatie over te brengen naar langetermijngeheugen in handzame pakketjes die 'schemata' worden genoemd. Deze schemata, c.q. de cognitieve representaties van complexe onderwerpen of constructen, kunnen vervolgens worden opgehaald uit het lange-termijngeheugen voor nader gebruik.

De ontwikkeling en verfijning van cognitieve schemata leidt tot leren vanuit cognitief perspectief. Complexere en meer verfijnde schemata zijn geassocieerd met meer expertise en automatisering. De activiteiten van het werkgeheugen bij elkaar wordt 'cognitieve belasting' of 'cognitive load' genoemd. Het werkgeheugen kan drie algemene activiteiten uitvoeren, te beschouwen als cognitieve belastings-subtypen. *Intrinsic Load* (intrinsieke belasting) treedt op wanneer lerenden essentiële taakelementen uitvoeren. Intrinsieke belasting is groter als taken complexer zijn of wanneer leerlingen minder ervaring hebben. Wanneer leerlingen werkgeheugencapciteit gebruiken om informatie-elementen vanuit de intrinsieke belasting combineren in leerschemata, vindt zogeheten *Germane Load* (leertaakgebonden belasting) plaats. Germane load is uiteindelijk wat leidt tot leren en automatiseren. *Extraneous Load* (irrelevante belasting) treedt op wanneer lerenden werkgeheugencapaciteit moeten gebruiken op een manier die niet productief is, zoals door externe of interne afleiding, of door onduidelijk of verwarrend onderwijs en instructie. Extraneous load draagt nooit bij aan leren. Wanneer de gecombineerde hoeveelheid intrinsieke, leertaakgebonden en irrelevante cognitieve belasting de werkgeheugen-capaciteit overschrijdt ontwikkelt zich cognitieve overbelasting die het leerproces en de prestaties negatief beïnvloedt. Volgens CLT is leren optimaal wanneer extraneous load wordt geminimaliseerd en intrinsic load wordt afgestemd op eerdere ervaring en kennis van de lerende; bij de gemotiveerde lerende bevordert dit leertaakgebonden belasting (germane load), wat leidt tot leren. CLT is veel bestudeerd in klassikaal onderwijs en experimentele proefopstellingen in de cognitieve psychologie, maar er is nu ook toenemend belangstelling voor de toepsassing in werkplaatsgebonden onderwijs in de gezondheidszorg.

CLT lijkt een ideale theoretische invalshoek om de opleiding van procedurele vaardigheden te bestuderen, omdat deze leerstof complex is (zoals hierboven beschreven) en er een aanzienlijk risico bestaat op cognitieve overbelasting. Tot nu toe is CLT in live vaardigheidsonderwijs nauwelijks onderzocht. Wij ontwierpen een serie studies in de veronderstelling dat de toepassing van CLT in live vaardigheidstaining van groot belang zou kunnen zijn, en hebben die in dit proefschrift beschreven. We kozen de gastro-intestinale (GI) endoscopie bij fellows in opleiding to gastro-enteroloog / endoscopist als een voorbeeld van een procedurele vaardigheid voor deze studies, omdat zij te beschouwen is als een "middle-of-the-road" -procedure in termen van complexiteit, tijdsduur en leerbaarheid, en vanwege haar essentiele betekenis voor de volksgezondheid. We hebben de studies ontworpen met het oog op generaliseerbaarheid voor een brede scala aan klinische vaardigheidstoepassingen.

Hoofdstuk 1 biedt een gedetailleerd overzicht van de uitdagingen van procedurele vaardighedentraining, theoretische principes van CLT, de redenen om CLT te gebruiken voor het bestuderen van procedurele vaardigheidstraining en om GI-endoscopie te bestuderen als een voorbeeld ervan.

Omdat cognitieve belasting zetelt in het werkgeheugen, moesten we een methode vinden om deze te meten of te schatten voor onze studies. De meest gebruikelijke aanpak is om eenvoudigweg mentale inspanning te schatten, maar wij vonden het essentieel om ook de subtypen cognitieve belasting (intrinsic, extraneous en germane load; IL, EL en GL) te meten, om theoretische en praktisch relevante conclusies te kunnen trekken. We zijn daarom begonnen met het proefschriftonderzoek door een psychometrisch instrument te ontwikkelen - de Cognitive Load Inventory for Colonoscopy (CLIC) - om intrinsic, extraneous en germane load te meten tijdens coloscopie-training (**Hoofdstuk 2**). We volgden een aanbevolen systematisch en rigoureus proces om deze *CLIC* te ontwikkelen en testte het vervolgens bij een representatieve, nationale steekproef van 477 fellows gastro-enterologie in de Verenigde Staten. **Hoofdstuk 2** beschrijft sterke steun voor de validiteit van de CLIC. We testten 19 items (8 voor IL, 6 voor EL, 5 voor GL) en behielden 15 in het uiteindelijke model (7 IL, 4 EL, 4 GL). Verkennende factoranalyse leverde een driefactoroplossing op, die goed paste bij de constructen van intrinsieke, irrelevante en leertaakgebonden belasting. Alle items pasten binnen het verwachte factormodel met uitstekende factorladingen. Confirmatory factoranalysis leverde bevredigende fitparameters op. **Hoofdstuk 2** concludeert dat de CLIC drie constructen meet bij lerenden die een coloscopie uitvoeren, die naar onze mening intrinsieke, irrelevante en leertaakgebonden cognitieve belasting representeren.

We hebben vervolgens gezocht naar factoren die verband houden met elk van de subtypen cognitieve belasting. Om dit te bereiken, hebben we in **Hoofdstuk 3** een groot aantal extra gegevens geanalyseerd van de 477 deelnemers in **Hoofdstuk 2**. We bevroegen deze deelnemers over meerdere kenmerken van zichzelf (de leerling), de taak/patiënt, de omgeving waar de verrichting was uitgevoerd en de supervisor. In multivariate modelvorming werden kenmerken geidentificeerd die verband houden met intrinsieke, irrelevante en leertaakgebonden belasting, en modellen ontwikkeld voor elk van de drie subtypen belasting, steeds theoretisch consistent met CLT. Intrinsieke belasting was voornamelijk geassocieerd met kenmerken van lerenden (negatieve associatie met jaar in opleiding en eerdere ervaring, positieve associatie met vermoeidheid) en patiënten/taken (positief associatie met de tolerantie die de patient vertoonde met de coloscopie en met aantal uitgevoerde verrichtingen). Irrelevante belasting was voornamelijk geassocieerd met kenmerken van de setting (positief met aantal coloscopieën eerder die dag) en de supervisor (negatief met supervisorvertrouwen en -betrokkenheid, positief met overname van de verrichting door de supervisor), en ook positief geassocieerd met moeheid van de lerende. Leertaakgebonden belasting was positief geassocieerd met supervisor-betrokkenheid. De verschillende patronen van factoren die met elk subtype van cognitieve belasting geassocieerd bleken, zijn theoretisch logisch en vormen aanvullende steun voor de valditeit van de CLIC. Terwijl het onderzoek in **Hoofdstuk 3** werd uitgevoerd voor training in coloscopie, zijn de onderzochte aspecten makkelijk te generaliseren naar andere live vaardigheidstraining.

Nadat we hadden geleerd cognitieve belasting te meten tijdens GI-endoscopietraining, en gewapend met een voorlopig inzicht in wat van invloed zou kunnen zijn op cognitieve belasting bij live vaardigheidstraining, kozen we ervoor ons te richten op de manier waarop specifieke onderwijsbenaderingen de cognitieve belasting zouden kunnen beïnvloeden. Hiertoe hebben we de relevante literatuur geëxploreerd, en een verkennend overzicht gemaakt van studies die het CLT-concept hebben gebruikt om werkplektraining van studenten in professionele beroepen te bestuderen, zowel binnen als buiten de gezondheidsberoepen (**Hoofdstuk 4**). Dit werd uitgevoerd als een zgn. Best Evidence Medicial Education (BEME) literatuurreview. Hoewel de geïncludeerde studies meestal een nauwe focus hadden en de subtypen cognitieve belasting vaak niet werden meegenomen, droegen de 116 uiteenlopende studies wel bij aan een aantal theoretische implicaties en praktische aanbevelingen van CLT bij vaardigheidstraining. Belangrijk is dat deze studies de de basis hebben gelegd voor de laatste twee hoofdstukken van het proefschrift.

Hoewel onderzoekers algemene onderwijsstrategieën hebben beschreven en onderzocht, gericht op de doelen van CLT, hebben maar weinig studies die in **Hoofdstuk 4** worden besproken aangegeven welke leerstrategieën kunnen worden gebruikt om de cognitieve belasting in live vaardigheidstraining in de gezondheidszorg te verbeteren. Wij wilden weten hoe experts dit onderwijs geven, en hebben daarom een kwalitatieve studie opgezet met 16 ervaren klinische docenten in de endoscopie in de Verenigde Staten, Canada en Nederland (**Hoofdstuk 5**). Via *purposeful sampling* werden deelnemers geselecteerd met goed inzicht in endoscopisch onderwijs en onderwijskundige kennis. We vroegen hen naar problemen die ze bij hun endoscopie-lerenden zagen en de leerstrategieën die ze gebruikten om hen te helpen deze uitdagingen te overwinnen. We analyseerden interview transcripten met *template analysis*. Deelnemers beschreven leeruitdagingen die we hebben gecategoriseerd als gerelateerd aan de lerende (*performance* versus *mastery* orientatie, lage self-efficacy, onvoldoende overzicht van wat men deed); aan de taak (psychomotorische uitdagingen, ontwikkeling van een mentaal model van de taak, tactiel gevoel en begrip); aan de leraar (relatie tussen docent en lerende, inadequaat onderwijs, veranderlijkheid in het lesgeven); en aan de setting (interne en externe afleidingen, systemische problemen). Deelnemers rapporteerden het gebruik van leerstrategieën die compatibel waren met CLT, inclusief die welke intrinsieke belasting afstemmen door te matchen met het niveau van de leerling (lesgeven langs een ontwikkelingscontinuüm, motorische instructie, technische assistentie), minimaliseren van irrelevante belasting (optimaliseren van de omgeving, systeem-oplossingen, emotionele ondersteuning, verwachtingen definiëren), en bevorderen van leertaakgebonden belasting (bevorderen van mastery, bevorderen van schemavorming, stop-en-focus strategie).

Ten slotte hebben we onderzocht hoe de leeractiviteiten uit deze interviews en andere in de praktijk worden uitgevoerd door 'typische' endoscopie docenten, en welke invloed lerenden ervoeren van deze strategieën op hun intrinsieke, leertaakgebonden en irrelevante cognitieve belasting. Om dit te onderzoeken (**Hoofdstuk 6**) voerden we een *mixed methods*-studie uit, waarbij we volledige coloscopieën observeerden die werden uitgevoerd door gastro-enterologische fellows (lerenden) en behandelende artsen (docenten) in twee grote ziekenhuizen. Van de onderwijsactiviteiten uit **Hoofdstuk 5** selecteerden wij er 11 die we geschikt achtten voor observatie: bewegingsinstructie, feedback, schema-teaching, checken van begrip, overnemen, verwachtingen stellen, technische assistentie, emotionele ondersteuning, omgevingsaanpassing, stimuleren van mastery doeloriëntatie, stoppen en luisteren. We gebruikten *event-sampling* en *real-time content-analyse* om te bepalen wanneer en hoe die 11 onderwijsactiviteiten plaatsvonden. Daarna werd elke fellow geïnterviewd om na te gaan welke invloed elke waargenomen activiteit had op de subtypen cognitieve belasting; de interviewtranscripten hiervan werden vervolgens via inhoudsanalyse onderzocht. We zochten met name het verband tussen activiteiten gericht op CL subtype leertaakbelasting met scores van de gemeten leertaakbelasting. De 11 onderwijsactiviteiten (interventies) werden in totaal 515 uitgevoerd bij 10 coloscopieën in 10 verschillende sessies bij 8 verschillende fellows. De onderwijsintensiteit varieerde sterk: van 0.7 tot 3.3 onderwijsactiviteiten per minuut proceduretijd. De perioden direct vóór en na de procedure werden zelden gebruikt voor onderwijs. Meest waargenomen bleken bewegingsinstructie, feedback, schema-teaching en checken van begrip. Er kwamen ook twee nieuwe onderwijsactiviteiten voor: cognitieve informatie verschaffing en autonomie geven. De meeste onderwijsactiviteiten werden gevoeld als vermindering van intrinsieke belasting en vergroting van leertaakgebonden belasting; nauwelijk als beinvloeding van irrelevante belasting. Maar soms veroorzaakten nuttige leeractiviteiten irrelevante cognitieve belasting, namelijk als de docent overdreven veel instructie gaf. Toepassing van activiteiten gericht op leertaakgebonden belasting correleerde matig positief met feitelijk gemeten leertaakgebonden belasting (Spearman's rho 0.44). We concluderen dat onderwijsactiviteiten kunnen worden gebruikt om doelen van CLT te bevorderen tijdens de training van verrichtingen, maar dat ze weloverwogen moeten worden toegepast om te voorkomen dat lerenden cognitief worden overbelast.

Het proefschrift eindigt meteen discussie in **Hoofdstuk 7**, waarin het werk in dit proefschrift wordt gecombineerd met andere literatuur over CLT en het trainen van vaardigheden. **Hoofdstuk 7** gaat specifiek in op enkele belangrijke bijdragen van dit werk aan de literatuur, waaronder: hoe subtypes van cognitieve belasting, conceptualisatie van leertaakgebonden belasting te meten? wat draagt bij aan intrinsieke, leertaakgebonden en irrelevante belasting in live vaardigheidstraining? en hoe onderwijs in te zetten voor een optimale balans van cognitieve belasting van lerenden die verrichtingen uitvoeren? Daarna volgen voorstellen voor toekomstig onderzoek, vooral in relatie tot CLT en onderzoek van gezondheidszorgonderwijs en gericht op het begrijpen van cognitieve overbelasting en op de relevantie van het meten van leren en presteren bij onderzoek van onderwijs in de gezondheidszorg en CLT. Ten slotte gaan we in op de sterke punten en beperkingen van de proefschriftstudies en pleiten we voor het generaliseren van de bevindingen naar diverse vaardigheidssettings.

Concluderend levert dit proefschrift een substantiële theoretische en praktische bijdrage aan ons begrip van onderwijs in live vaardigheidstraining voor beroepen in de gezondheidszorg via het perspectief van CLT. We hopen dat dit een stimulans is voor de ontwikkeling van onderzoek, onderwijs en curriculuminnovatie, met het doel het leren te optimaliseren op dit uitdagende, maar interessante terrein in de gezondheidszorg.

Acknowledgements

The opportunity to complete this body of work as a practicing academic gastroenterologist has truly been a privilege. Though planning and hard work have contributed much to my professional successes, serendipity has played a role as well, and I am grateful to the people who have shown up in my life and helped me to reach this point in my career.

I am incredibly grateful to my amazing doctoral team: Drs. Olle ten Cate, Patricia O'Sullivan and Christy Boscardin. They all have such amazing knowledge and insight into education research, yet it has been their commitment to my growth and development that has been of greatest significance to me. Throughout this process, they have managed to be demanding and rigorous, yet at the same time supportive and encouraging. I have always been able to count on them to be on my side, to be my advocates, and to promote my career development. I am especially grateful to Pat for her tireless dedication to helping me complete my research, especially for serving as the second reviewer for the more than 4,000 titles in our systematic review! Thank you all for believing in me and for connecting me to the health professions education community of practice at UCSF and throughout the United States and the Netherlands. Thank you for always pushing me to excel and succeed in this field. And thank you for providing a wise perspective and words of reassurance to help me manage the inevitable obstacles that we encountered along the way.

I thank my division chief, Jackie Maher, who has supported my education aspirations since the beginning, including covering my effort to attend Teaching Scholars for an entire year, paying for my doctoral tuition and publicizing my efforts at division meetings.

I thank my doctoral cohort colleagues: Judy Bowen and Jody Steinauer and Leslie Floren. It has been a joy to get to know each of you. I so appreciate your support and counsel and friendship, and I am excited about each of your lines of research and career paths!

I thank John Q. Young, my cognitive load brother. I love discussing cognitive load theory with you, and I am excited about our current and future collaborations.

I thank the others coauthors outside of my doctoral team who contributed to individual studies: Tamara van Gog, Lauren Maggio, Judy Bowen, and Brijen Shah. Your contributions strengthened those papers significantly!

I thank cognitive load theory experts in the Netherlands: Jeroen van Merriënboer, Tamara van Gog and Jimmie Leppink, who have been so gracious with their time and expertise. I hope to work with each of you further in the future.

I thank the Jacobsohn Endowment for Excellence at Zuckerberg San Francisco General Hospital for the grant funding that made my first two doctoral studies possible, and Steve Jacobsohn for being a dedicated colleague who always asks how things are going with my research.

I thank the Society of Directors of Research in Medical Education for grant funding to help support my systematic review, and also the Best Evidence in Medical Education staff and reviewers for guidance and assistance in completing that behemoth effort.

And of course, I cannot adequately express the gratitude I have for my family. Hazen, your support for my professional pursuits throughout our life together has always been a source of strength for me. Thank you for encouraging me through the dark times and celebrating with me during the bright times. You are my soulmate forever, and I would not be where I am without you! To my children – Grace, Lucas and Eleanor – thank you for being the joys of my life. Your love, silliness and effervescence bring a smile to my heart every day. And to Grace I add an additional shout-out for being able to explain cognitive load theory better than anyone I know outside of my research team!

Curriculum Vitae

Justin Louis Sewell was born in the State of Texas in the United States and grew up across the southern United States. He spent his adolescence in Tucson, Arizona, and attended the University of Arizona, where he earned his bachelor's degree in molecular and cellular biology, followed by a master's degree in publish health, and then his medical degree. After completing medical school, Justin moved with his wife Hazen to San Francisco to complete a residency in internal medicine followed by a fellowship in gastroenterology at the University of California San Francisco (UCSF). Upon completing his medical training, he joined the faculty at UCSF in the Department of Medicine and Division of Gastroenterology. Shortly after joining the faculty, he sought training in education through the UCSF Teaching Scholars Program. Although he had substantial experience designing and performing clinical and health services research, through Teaching Scholars, he found his academic calling in education research. To further promote the nascent education research skills developed in Teaching Scholars, Justin joined the UCSF – Utrecht University collaborative doctoral program in health professions education in 2014.

Justin has had a faculty appointment at UCSF for 8 years. He is currently an Associate Professor of Medicine and has an academic career focused on health professions education. In addition to education research, he is involved in direct teaching across the spectrum of health professions learners, curriculum design, education leadership, and mentorship of students, residents, fellows and faculty on scholarly projects. Justin is a course director for a major first-year medical school course and serves on multiple education committees. He led development of two primary components UCSF's recently implemented curriculum (UCSF Bridges Curriculum), which is a fully integrated four-year medical school curriculum with many innovative elements. He continues to practice as a gastroenterologist seeing patients and performing gastrointestinal endoscopic procedures (and trying to not cognitively overload his trainees) at the Zuckerberg San Francisco General Hospital. He lives in San Francisco with his wife Hazen, and three children Grace, Lucas and Eleanor.

SELECTED PUBLICATIONS

1. **Sewell JL**, Maggio LA, ten Cate O, van Gog T, Young JQ, O'Sullivan PS. Cognitive load theory for training health professionals in the workplace: A BEME review of studies

among diverse professions: BEME Guide No. 53. *Med Teach*. 2018 Oct 17; 1-15 [Epub ahead of print].
2. Pourmand K, **Sewell JL**, Shah BJ. What Makes a Good Endoscopic Teacher: A Qualitative Analysis. *J Surg Educ*. 2018 75(5):1195-1199.
3. **Sewell JL**, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. 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*. 2017 Nov; 92(11):1622-1631.
4. Dilly CK, **Sewell JL**. How to Give Feedback During Endoscopy Training. *Gastroenterology*. 2017 Sep; 153(3):632-636.
5. Young JQ, Boscardin CK, van Dijk SM, Abdullah R, Irby DM, **Sewell JL**, ten Cate O, O'Sullivan PS. Performance of a cognitive load inventory during simulated handoffs: Evidence for validity. *SAGE Open Med*. 2016; 4:2050312116682254.
6. **Sewell JL**, Maggio LA, Artino AR. Planning the Literature Review. *Acad Med*. 2016 Dec; 91(12):e18.
7. Maggio LA, **Sewell JL**, Artino AR. The Literature Review: A Foundation for High-Quality Medical Education Research. *J Grad Med Educ*. 2016 Jul; 8(3):297-303.
8. **Sewell JL**, Boscardin CK, Young JQ, ten Cate O, O'Sullivan PS. Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. *Med Educ*. 2016 Jun; 50(6):682-92.
9. Young JQ, **Sewell JL**. Applying cognitive load theory to medical education: construct and measurement challenges. *Perspect Med Educ*. 2015 Jun; 4(3):107-9.
10. **Sewell JL**, Day LW, Tuot DS, Alvarez R, Yu A, Chen AH. A brief, low-cost intervention improves the quality of ambulatory gastroenterology consultation notes. *Am J Med*. 2013 Aug; 126(8):732-8.