CARDIAC AND RESPIRATORY AUSCULTATION SKILL DEVELOPMENT
DOCTOR OF PHILOSOPHY (2013)  McMaster University
(FACULTY OF HEALTH SCIENCES – Hamilton, Ontario
Health Research Methodology)

TITLE:  Cardiac and respiratory auscultation skill development: exploration and application of cognitive load theory in health professions education

AUTHOR:  Ruth Chen B.A. (University of California, Berkeley), M.S.N. (Yale University)

SUPERVISOR:  Geoffrey R. Norman PhD

NUMBER OF PAGES:  ix, 106
ABSTRACT

This thesis explores the Cognitive Load Theory (CLT) framework with health professions education research, and applies the principles of CLT to one specific area of health professions education: the acquisition of cardiac and respiratory auscultation and physical assessment skill in undergraduate nursing students. The first study evaluates context-based learning environments on the acquisition of auscultation skill and physical assessment performance. Results suggest that for novice level students, high-context-based learning environments may contribute to extraneous cognitive load and may not be beneficial for learning. The next cluster of studies evaluates auscultation test performance following manipulation of cognitive load variables. The interleaving approach instruction was found to be most helpful for auscultation test performance. The series of studies conducted in this thesis demonstrate a useful direction for health professions research and promotes the use of cognitive load theory as a framework for instructional design and evaluation.
ACKNOWLEDGMENTS

My sincerest thanks to the entire committee for all the guidance, support, and feedback you’ve provided for this thesis. To my supervisor, Geoff Norman, and to committee members, Lawrence Grierson, Kelly Dore, and Rose Hatala, you all know that this thesis is a result of your dedication and input over the past few months. I am grateful to have you all as colleagues and mentors. I can only hope that I will develop even half the insight and understanding you all have as I continue to grow in my research abilities.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 1: Cognitive Load Theory: Implications for Nursing Education and Research</td>
<td>3</td>
</tr>
<tr>
<td>References</td>
<td>21</td>
</tr>
<tr>
<td>Chapter 2: Exploration of Context-Based and Non-Context-Based Instruction on Auscultation Skill Development</td>
<td>27</td>
</tr>
<tr>
<td>Research Questions</td>
<td>31</td>
</tr>
<tr>
<td>Methods</td>
<td>32</td>
</tr>
<tr>
<td>Results</td>
<td>39</td>
</tr>
<tr>
<td>Discussion</td>
<td>45</td>
</tr>
<tr>
<td>Conclusion</td>
<td>50</td>
</tr>
<tr>
<td>References</td>
<td>51</td>
</tr>
<tr>
<td>Chapter 3: Manipulation of Cognitive Load Variables and Impact on Auscultation Test Performance</td>
<td>57</td>
</tr>
<tr>
<td>Study 3.1. Modification of intrinsic cognitive load</td>
<td>61</td>
</tr>
<tr>
<td>Research Question</td>
<td>63</td>
</tr>
<tr>
<td>Methods</td>
<td>63</td>
</tr>
<tr>
<td>Results</td>
<td>65</td>
</tr>
<tr>
<td>Discussion</td>
<td>68</td>
</tr>
<tr>
<td>Study 3.2. Facilitating germane load through multiple representations</td>
<td>70</td>
</tr>
</tbody>
</table>
Lists of Figures and Tables

Tables

Chapter 2

Table 1. Total Auscultation Test Scores and Cardiac and Respiratory Test Subscores Between Instruction Groups .............................................................. 39

Table 2. Auscultation Test Subscores for OLD Total, Cardiac, and Respiratory Sounds .............................................................. 41

Table 3. Auscultation Test Subscores for NEW Total, Cardiac, and Respiratory Sounds .............................................................. 42

Table 4. Clinical Simulation Performance Scores Between Instruction Groups ..... 44

Chapter 3

Table 1. Post – Instruction Auscultation Test Scores, Cardiac and Respiratory Test Subscores, and Respiratory Retest Scores Between Instruction Groups ……… 66

Table 2. Auscultation Test Scores, Cardiac and Respiratory Test Subscores, and Respiratory Retest Scores Between Instruction Groups ………………… 75

Table 3. Auscultation Test Scores, Cardiac and Respiratory Test Subscores, and Respiratory Retest Scores Between Instruction Groups ……………… 81

Appendix

Table 1. Descriptive data from scale, averaged across observers ...................... 103
Figures

Chapter 1

Figure 1. Cognitive load variables  ................................................................. 8

Chapter 2

Figure 1. Study Design  ....................................................................................... 32

Figure 2. Auscultation Test Scores for Instruction Groups.  ............................ 40

Chapter 3

Figure 1. Auscultation Test performance on OLD and NEW Respiratory sounds  ... 68

Figure 2. Interaction between Mixed and Blocked instruction with Auscultation Test performance on Recall (OLD) and Near Transfer (NEW)  ......................... 82

Appendix

Figure 1. Physical Assessment Measurement Scale  ........................................... 100
DECLARATION OF ACADEMIC ACHIEVEMENT

I, Ruth Chen, declare that this document is my own and that I am the sole author of its contents. I declare that this is a true copy of my dissertation, including any final revisions, as approved by my supervisor and supervisory committee and that this dissertation has not been submitted to any other university or institution.
Introduction

An educator’s responsibility is to teach so students will learn. However, approaches to address seemingly straightforward questions such as what should be taught, how and in what contexts should that material be taught, and how should that learning be measured are often unclear. One assumes that if an educator merely teaches what he knows, then the student will also simply learn what the educator believes the student should know.

However, in health professions education, it is not enough to say that if one has practiced for many years as a health professional, then one will also have the requisite knowledge and ability to teach the professional practice effectively to others. An understanding of principles of learning, working memory, and cognition, as well as exposure to instructional manipulations that might further elucidate learning processes is required. Therefore, it is from this perspective that the following thesis was written. This thesis will explore the work conducted in the field of cognitive psychology, with a focus on Cognitive Load Theory, in an attempt glean principles of learning and instruction, with the intention of applying these principles to one specific area of nursing education research: the acquisition of cardiac and respiratory auscultation and physical assessment skill in undergraduate nursing students.

Chapter 1 provides an overview to Cognitive Load Theory and discusses the impact of manipulating intrinsic, extraneous, and germane load on instruction and learning. This literature review also discusses applications of CLT for health
professions education and research. A scenario from nursing education in pediatric cardiology is provided as an example for how CLT can be applied to instructional design.

**Chapter 2** focuses on context-based learning environments and the use of simulation technologies to facilitate authentic learning experiences. While there is theoretical rationale for placing learners in high context-based, or authentic, environments as it facilitates emotional engagement, the study conducted in this chapter explores the potential for highly authentic environments to contribute extraneous cognitive load, resulting in ineffective learning and cognitive overload. Thus, the potential drawbacks of high context-based learning environments resulting in extraneous cognitive load imposed on learners’ working memory are explored. The **Appendix** provides details regarding the development and psychometric testing of the measurement scale that was used as one of the outcome measures in this chapter.

**Chapter 3** evaluates the effect of manipulating cognitive load variables on learners’ auscultation test performance. Through a series of studies, the intrinsic, extraneous, and germane loads of auscultation skill instructional design are examined. Strategies used in the literature to facilitate learning – the use of multiple representations and interleaving approaches – will be the focus in this chapter.

Finally, **Chapter 4** will provide a general summary of the studies conducted in this thesis. Future directions for health professions education research based on the findings from these studies will be discussed.
Chapter 1

Cognitive Load Theory: Implications for Nursing Education and Research

Students in Bachelor of Science in Nursing (BScN) programs acquire a body of knowledge, skills, and attitudes during their undergraduate education in preparation for future practice. One goal of an educator, therefore, is to create an environment of instruction where this learning could take place. The purpose of this review is to provide an overview of Cognitive Load Theory (CLT) and the impact that cognitive load has on working memory and learning. This review will also explore how work on CLT in the field of cognitive psychology has been applied to health professions education research, particularly for multimedia and simulation-based applications. Finally, the implications of CLT on nursing education and research will be discussed.

Working Memory and Learning

For the purposes of this review, an operational definition of learning is the student’s acquisition of knowledge, skills, or attitudes (Van Merriënboer, Kirschner, & Kester, 2003) that results in changes to long-term memory (Norman, 2013) and which produces an observable knowledge, behavior, or action outcome. Central to this definition of learning is that a student receives information through multiple sensory pathways (e.g. visual, auditory inputs through pictures, words, sounds), and creates visual and auditory representations within the cognitive system (Mayer, 2002). These representations are processed within the structures of working memory with the goal of transferring and storing this information into long-term memory.

Baddeley’s model of working memory describes the cognitive systems required for students to process complex tasks such as reasoning, understanding, and learning.
(Baddeley, 2010; Burgess & Hitch, 1999). In this model, there are three components of working memory that serve to facilitate information processing. The central executive component, responsible for organizing knowledge and information, interacts with two short-term memory storage components: the visual-spatial sketch pad and the phonological loop, with the visual-spatial sketch pad responsible for processing visual information and the phonological loop for processing verbal-acoustic material (Baddeley, 1992). Together, the central executive, the visual-spatial sketch pad, and the phonological loop comprise the working memory structures which interact with, and produce changes to, long-term memory (Mayer, 2002).

**Cognitive Load Theory**

There are several assumptions about the cognitive architecture of working memory and long-term memory. The first assumption is that working memory is constrained and limited. Our understanding around limited working memory was first detailed by Miller who stated that an individual was only capable of retaining “seven plus or minus two” units of information at any point in time (Miller, 1956). Therefore, any greater amounts of information presented that exceed the capacities of a learner’s working memory could not be retained. The second assumption is that there is virtually unlimited long-term memory; furthermore, working memory and long-term memory structures can interact. Therefore, inasmuch as information processing within working memory could be retained in the infinite stores of long-term memory, information could likewise be brought forth from long-term memory to interact with and facilitate working memory processes (Schnotz & Kürschner, 2007). One final assumption is that the cognitive load imposed on a learner’s working memory during instruction can be
modulated. Thus, the student’s cognitive load could be increased or decreased, and processing facilitated or hindered (Mousavi, Low, & Sweller, 1995).

Our understanding of Cognitive Load Theory (CLT) is historically rooted in John Sweller’s work in understanding learners’ problem solving strategies (Sweller, 1988). In one problem solving approach, means-ends analysis, students identify a problem and an intended goal and work out strategies to minimize the gap between the problem state and the goal state by holding within working memory all the operations required to move from one step to another, the sub-goals, and the relationship between these sub-goals. This approach is taxing for working memory and results in a high cognitive load. Sweller also noted differences in problem solving approaches between experts and novices; novices would more frequently employ the above means-ends analysis approaches, whereas experts would use a cognitively more efficient process moving forward from problem state to intended goal.

Sweller first termed the cognitive resources required for complex problem solving as “cognitive processing capacity” and he argued that the cognitive load imposed on a learner during problem solving could potentially interfere with learning (Sweller, 1988). In other words, the actual cognitive work required to figure out how to solve a problem could interfere with a learner’s ability to learn the actual principles the problem intended to teach. Given a learner’s limited working memory, it would be helpful for nurse educators to understand the principles of CLT that can impact student learning.

**Three components of cognitive load.** There are three components of cognitive load: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Van Merriënboer & Sweller, 2005). Intrinsic cognitive load describes the actual learning
goal or task. The intrinsic cognitive load is directly related to the amount of information required to learn. Along with the learning goal or units of information provided in instruction, intrinsic cognitive load also considers the inherent difficulty of the learning goal and the level of expertise regarding the subject matter by the learner. One measure of intrinsic cognitive load is element interactivity which describes the number of separate components in the learning goal that would need to be held simultaneously in working memory to process the presented information (Leahy & Sweller, 2005). The greater the complexity of the learning goal, the greater the intrinsic cognitive load (Paas, Renkl, & Sweller, 2003).

In addition to the intrinsic cognitive load imposed by the learning goal, the format and features of instruction can contribute additional extraneous cognitive load (Sweller, Van Merrienboer, & Paas, 1998). Extraneous cognitive load is attributed to features of instruction that are not necessary for learning and which therefore impose a burden on working memory’s cognitive processing ability. CLT assumes an additive model for the intrinsic and extraneous cognitive load variables such that, for any particular learning task or goal, the sum of the intrinsic and extraneous load must not exceed working memory capacity (Sweller, 1994). Figure 1 below depicts the additive nature of the cognitive load variables.

An additional variable added later to CLT was the concept of germane cognitive load. Germane load describes the processing that improves automation of information into long-term memory and which promotes learning (Paas & Van Merriënboer, 1994). Germane load has been described elsewhere as “generative cognitive processing” (Mayer, 2010) wherein the learner is better able to “make sense of” and understand the
presented material. Germaine load, while increasing the overall cognitive load of the learner was identified as distinct from extraneous load in that the instruction approaches promoted, rather than detracted from, understanding of the instructional material. Germaine cognitive load is also considered as an additive variable such that any working memory available beyond the required intrinsic cognitive load could be enhanced by germaine cognitive load. Therefore, in applying the principles of germaine cognitive load to the CLT framework, nurse educators must seek to keep intrinsic cognitive load within the limits of working memory, decrease extraneous cognitive load within the instructional design, and finally, facilitate processing of information by promoting germaine cognitive load.

Identifying the appropriate amount and type of cognitive load imposed on a learner during instruction is a significant factor that determines the success of an educator’s instructional intervention (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Poorly designed learning goals that require complex processing of multiple ideas or skills and which exceed the capacity of a learner’s working memory lead to cognitive overload and decreased learning (Doolittle, McNeill, Terry, & Scheer, 2005). As such, either too much learning material or material that is too complex (intrinsic load) or poorly designed instruction that includes unnecessary information or instructional features (extraneous load), or a combination of both, can lead to an overload to working memory (see Figure 1, part (a)). Instructional design can be optimized to decrease extraneous cognitive load (part (b) in Figure 1) such that the additional working memory capacity could be used for germaine cognitive processing (part (c) in Figure 1). As nurse educators,
understanding how to optimize student learning and success by tailoring our instruction based on an understanding of cognitive load variables would be worthwhile.

Given the nature of the three components of cognitive load, instructional materials can be constructed for nursing students that optimize the cognitive load imposed on working memory and facilitates retention in long-term memory. To provide an example to describe how CLT might be incorporated into the design of nursing instructional goals, two characters will be threaded throughout the following sections. Clinical Professor “AA” is a pediatric nurse who teaches a module on pediatric congenital heart disease and physical assessment. Senior-level nursing student “ZZ” represents a typical student in the course, and Professor AA wishes to use principles of CLT to inform the instructional design of the learning module.

**Approaches to adjust intrinsic cognitive load.** The first variable, intrinsic cognitive load, describes the inherent nature of the learning task at hand. Therefore,
modifying intrinsic cognitive load necessarily involves changing the learning task by either adding or subtracting the amount and complexity of the material to be learned. In our scenario, Professor AA wishes to provide instruction on the management of congestive heart failure (CHF) of a pediatric patient with a cardiac anomaly. This is a nursing process that possesses high intrinsic cognitive load due to high element interactivity (e.g. requiring an understanding of cardiac and respiratory physiology, changes in patient presentation, nursing assessments and actions, and the interconnectedness between the variables), the novice level of expertise by Student ZZ (who would have had minimal prior exposure to pediatric cardiac management over the course of a general undergraduate nursing program), and the high inherent task difficulty (learning how to manage a child with CHF is a more complex task than being able to verbalize the signs and symptoms of CHF; this is still more complex than being asked to describe the path of blood flow through the normal heart structures).

Therefore, modifying intrinsic cognitive load necessarily involves changing the nature of the learning goal or task such that the goal is simplified or made more complex based on the level of expertise of Student ZZ and the inherent difficulty of the learning goal itself. If a learning goal remains unchanged, the intrinsic cognitive load is otherwise considered unchangeable. As nurse educators, structuring learning goals that keep the intrinsic cognitive load within the limits of a student’s working memory helps promote efficient learning and prevents cognitive overload.

**Approaches to minimize extraneous cognitive load.** The cognitive and educational psychology literature contains many studies of strategies to minimize extraneous cognitive load. Fortunately, even a basic understanding of these strategies can
yield significant applications for nursing education and research. This review will not provide a comprehensive overview of all strategies that have been explored to minimize extraneous load, but a few strategies will be highlighted.

The Worked Example Effect was one of the first reported strategies. These studies compared students who were provided with explicit details regarding the steps necessary to solve a problem with students who were not provided with these details and were therefore required to figure out what those steps might be on their own (Sweller et al., 1998). Through seeing worked examples, students conserved cognitive resources which allowed them to focus on the particular learning goal in the instructional session rather than expending cognitive resources by attempting to solve the problem in an unsystematic or ‘trial and error’ manner.

Therefore, for Professor AA, when teaching about management of CHF in a pediatric patient, worked examples might come in the form of providing students with key cardiac and respiratory assessment measures and describing how changes in physical findings might suggest deterioration or improvement in patient status rather than asking students to generate a management plan without any additional instructional support. With this approach, the goal of helping Student ZZ understand CHF management (the intrinsic load) is facilitated by decreasing the extraneous load of Student ZZ, who might not understand all the components involved in managing (or ‘solving’) this patient’s CHF issue (or ‘problem’).

Another approach to minimize extraneous cognitive load is the Split Attention Principle (Ayres & Sweller, 2005). Studies conducted by Ayres, Sweller, and others have shown that students who were required to focus on multiple disparate objects at once (e.g.
a diagram and text description that were separated on a page) experienced increased extraneous cognitive load compared to students who focused on objects that were integrated (e.g. a text description placed next to the appropriate part of a diagram) (Chandler & Sweller, 1991). This principle was applicable not only for visual information, but also applied when there was competing auditory information during instruction. In one study, learning and retention significantly decreased when students received instruction that included narration and background music that accompanied the narration in comparison to students who received the narration without background music (Moreno & Mayer, 2001). Therefore, for Student ZZ learning about the physical findings of pediatric patients with CHF, the split attention principle suggests that removing competing or non-integrated visual stimuli (e.g. simultaneously needing to view data from a cardiac monitor and assess the patient for tachypnea and retractions) or removing competing auditory information (e.g. listening for the patient’s heart rate but also hearing the cardiac monitor alarm bells ringing) would result in improved learning.

A third strategy to reduce extraneous cognitive load is the Modality Principle (Low & Sweller, 2005). Even as working memory is limited, the learner’s visual-spatial sketchpad (visual input) and phonological loop (auditory pathway) within working memory can work synergistically to process instructional material. When using multimedia for instruction, for example, if a diagram has accompanying text, educators can decrease extraneous cognitive load by converting the text into an auditory narration while maintaining the diagram in the visual format. Through the modality principle, instructional design is optimized when visual and auditory processing pathways can both be engaged.
Approaches to foster germane load. Germane load is the cognitive load resulting from activity in working memory that facilitates learning beyond simple task performance (Schnotz & Kürschner, 2007). While extraneous cognitive load interferes with learning by unproductively overtaxing working memory, germane cognitive load promotes acquisition and automation of information into long-term memory (Paas et al., 2003). Therefore, efforts both to decrease extraneous load and to increase germane load during instruction is advocated, with the goal that the total cognitive load does not exceed working memory capacity (Van Merriënboer & Sweller, 2005). Paas and Van Merrienboer (1994) discuss the Variability Effect, whereby increases in the variability of learning tasks may contribute to increased cognitive load but resulted in improved learning outcomes. This would seem to contradict the previous examples highlighting the negative effects of increased extraneous cognitive load on learning. Indeed, the initial explorations of CLT variables first focused on intrinsic cognitive load and strategies to decrease extraneous cognitive load. The concept of germane load was introduced later when researchers discovered that some forms of instruction that ostensibly increased cognitive load were found to be beneficial for learning.

Another strategy to foster germane learning processes was to encourage students to make active comparisons and to articulate the differences across examples from different categories (Gerjets, Scheiter, & Schuh, 2008). This approach, while increasing cognitive demands from the learner, instead facilitated, rather than distracted from, learning. In another study, both providing worked examples (to decrease extraneous cognitive load) and prompting students to identify underlying principles illustrated by the examples of instruction (to enhance germane load) were explored. In this study, worked
examples were gradually faded out as the learner improved in understanding of the instructional materials, and students were then encouraged to articulate what the underlying principles were for the worked examples (Atkinson, Renkl, & Merrill, 2003). This instructional approach was found to improve learning and transfer significantly. The results have been consistent with other studies that attempt to facilitate germane cognitive load by asking students to provide self-explanations of the principles highlighted in the instructional material (Chi, Bassok, Lewis, Reimann, & Glaser, 1989).

Defining what constitutes extraneous cognitive load versus germane cognitive load can be difficult. Depending on the learning goal, the instructional design, and the level of learner expertise, factors contributing to extraneous load in one group of learners may serve to facilitate germane load in another group (Paas, Renkl, & Sweller, 2004). Therefore, instructional strategies that increase the cognitive load on working memory may end up contributing extraneous or germane load. Determinations regarding which form of cognitive load was attributed to the instructional design may at times be made post hoc (de Jong, 2010).

For Professor AA instructing students on the management of pediatric CHF, therefore, a potential strategy to facilitate germane load could be to provide clinical examples of patients displaying signs of CHF. Student ZZ would be asked to explain how a moderately sized ventricular septal defect (VSD) might contribute to an infant’s tachypnea. This approach, however, would presuppose Student ZZ understood the fundamental principles of cardiac and respiratory physiology in an infant without a cardiac anomaly and that the intrinsic load of this learning goal (CHF signs and symptoms in an infant with a VSD) did not already exceed the limits of the Student ZZ’s
working memory.

**Description of how cognitive load variables were measured.** Sweller stated that a direct measure of cognitive load imposed by a particular procedure or strategy was not available; however, any measure must account for the difficulty of the problem, the learner’s knowledge of the subject, and the instructional strategies employed (Sweller, 1988). Therefore, Sweller proposed a few strategies to measure cognitive load variables. Intrinsic cognitive load could be correlated with the number of units of information a learner was required to hold in working memory, with the general rule that an increase in instructional content resulted in increased cognitive load.

Another strategy was to evaluate the complexity of the instructional material whereby a learner needing to decide what to do next in a sequence of steps or a learner presented with greater number of possible options from which to choose would shoulder a greater cognitive load burden than when that learner was presented with instructional tasks that were less complex. An example often provided in the cognitive psychology literature is the task of learning of a new language. A greater number of vocabulary words to be learned would understandably translate into increased intrinsic cognitive load when compared with learning fewer vocabulary words. However, in addition to actual units of information measured, the complexity of the learning task would also contribute to increased cognitive load. Therefore, learning grammatical structure of a new language necessarily imposes a greater cognitive load than learning vocabulary words in isolation because learning grammar requires the ability to know vocabulary words as well as grammatical rules and syntax. Other ways to measure cognitive load would include
number of cycles required to reach solution or the number of sub-goals contained in the problem (Sweller, 1988).

Schnotz and Kurschner (2007) also state that there is no definitive way to measure cognitive load variables beyond the general estimating approaches employed in the literature. Three strategies the authors highlight include asking learners for subjective ratings of perceived cognitive load, measuring physiologic parameters, and applying performance-based measures. All of these approaches are limited in that each of these measures cannot distinguish between the different forms of cognitive load. Therefore, instructional approaches using CLT and research involving the manipulation of cognitive load variables will necessarily require a theoretical estimation of the cognitive load imposed on a learner’s working memory that may not be based on a direct or quantifiable measurement.

Elaboration of the three approaches described by Schnotz and Kurschner to measure cognitive load is beyond the scope of this review, but for the purposes of nursing education and research, it is helpful to be aware that there are no precise measures of cognitive load; however, research in nursing education using CLT as a framework of instruction can still include an identification of which variables are being manipulated, and can be investigated experimentally in a manner consistent with the hypothesis testing approaches of evidence-based nursing.

Current Applications of Cognitive Load Theory (CLT) to Education Research

Application of CLT to health professions education research. Principles of Cognitive Load Theory and its applications to health professions education research have been discussed with increasing frequency over the past several years, especially with
Multimedia learning in health professions education includes any approach that incorporates visual, auditory, and/or multi-sensory experience(s) into the instructional design. Therefore, with the increase in the applications of multimedia and other learning technologies such as high-fidelity simulation, there has been a concomitant increase in discussion regarding how these learning aids might impact on cognitive load and working memory. The literature in health professions education research include studies that call for the application of CLT principles to instructional design, studies that have used CLT as the framework for instructional design, and review articles providing support for how to apply CLT to health professions education. Furthermore, CLT is being used as a framework of instruction in comparison to other learning frameworks used in health professions education.

The literature calling for the application of cognitive load principles to instructional design include studies with virtual patients and the use of computer animations in medical education (Cook, 2009; Ruiz, Cook, & Levinson, 2009). In these articles, the authors highlight the need to evaluate technology- and multimedia- based instruction from the perspective that these more technologically sophisticated instructional formats might hinder learning by placing increased extraneous cognitive load on the learner. These authors challenge an unquestioning uptake of learning technologies and multimedia applications that do not adequately consider the instructional aims, the learner characteristics, and the evaluation metrics from a cognitive load perspective.
Other health professions education studies have applied the principles of CLT to
the instructional design (Holzinger, Kickmeier-Rust, Wassertheurer, & Hessinger, 2009;
Stark, Kopp, & Fischer, 2011) and highlight how CLT offers a useful framework for
evaluating the effectiveness of instructional designs and approaches. Holzinger et al
(2009) evaluated the effect of a hemodynamics simulation instructional group in
comparison with text-based instruction group and a group who received the simulation
instruction with additional feedback and support. The learning benefit occurred only
when additional instructional feedback and support was provided to the simulation-based
learning group; otherwise, the simulation group did not demonstrate improved learning
outcomes in comparison with the text-based instruction group. Because a fourth
intervention group was not included, text based instruction with instructional support, it
was difficult to ascertain whether these same learning benefits would persist if the
additional support were provided to students in the text-based learning environment as
well. The authors interpreted the findings through the lens of CLT and interpreted the
simulation-only instruction format as resulting in cognitive overload, but that providing
the additional support of instruction with the simulations facilitated learning and
processing of hemodynamics instruction. Further interpreted within the vocabulary of
CLT, simulation-only instruction could be seen as contributing to excessive extraneous
cognitive load, but providing the additional feedback and support during instruction
facilitated germane processing.

Stark et al (2011) explored the variables of example format and feedback
approach in managing cognitive load of medical students receiving hypertension and
hyperthyroidism instruction. Results suggested that offering students erroneous examples
with feedback that elaborated on the correct responses improved student performance, whereas erroneous examples without the elaborated feedback resulted in decreased performance. Furthermore, there was a greater positive learning effect of elaboration feedback when the difficulty of the learning domain increased (i.e. from instruction in hypertension to instruction on hyperthyroidism). Such studies demonstrate the useful ways that CLT, and evaluating instructional design with the goal of minimizing extraneous cognitive load and facilitating germane load, can contribute to nursing education and research.

Several review studies have highlighted the contributions that CLT-based approaches can make to health professions education, from general overviews discussing CLT as a potential framework (Patel, Yoskowitz, & Arocha, 2009; Rikers, Van Gerven, & Schmidt, 2004; Valcke & De Wever, 2006), to specific applications in the development of anatomy animations in medical education (Khalil, Paas, Johnson, & Payer, 2005a, 2005b). Discussion of CLT in the design of instructional material in health professions education includes strategies that reduce extraneous cognitive load, facilitate germane cognitive load, and incorporate learner expertise into instruction approaches (Van Merriënboer & Sweller, 2010).

Lastly, comparing the learning outcomes from two instructional frameworks – cognitive load theory and authentic, or context-based, learning – has been explored. The principle of authentic or context-based learning focuses on student engagement in ‘real’ environments as the key predictor of effective learning whereas cognitive load theory focuses on instructional design that decreases extraneous cognitive load and facilitates germane load as central to effective learning. La Rochelle et al (2011) created
instructional formats representing varying degrees of authenticity of the learning context for medical student instruction on polyuria, abdominal pain and anemia content. The expectation was that increased authenticity, or context-based learning, would result in improved subsequent clinical reasoning performance. While students found the authentic learning contexts more emotionally engaging, there was no significant difference in performance outcomes between those students randomized to receive paper-based instruction (lowest in authenticity, but with decreased extraneous cognitive load), DVD-based instruction, or standardized patient-based instruction (highest in authenticity, but with greatest perceived extraneous cognitive load).

**Applications of CLT to undergraduate nursing education and research.**

Currently, there are no known studies employing Cognitive Load Theory as framework for the design and evaluation of nursing education and research. General areas such as cognition and decision-making in nursing practice are explored and many other frameworks have been discussed in the literature.

For example, high-fidelity simulation and other multimedia applications are being incorporated into nursing education with enthusiasm and rapid acceptance. Frameworks used to guide the development of simulation-based learning experiences include behavioral, constructivist, and experiential learning approaches (Kaakinen & Arwood, 2009) and focus on instructional design within a paradigm similar to La Rochelle et al’s authenticity- or context- based approach. The underlying premise of these approaches is that students learn best when placed in authentic learning environments that most closely approximate actual clinical settings because it increases learner motivation and emotional engagement. As seen in the study by La Rochelle et al (2011), evaluating these
instructional modalities through the lens of the cognitive load imposed on learners would allow educators to reconsider how best to use these applications and modalities in nursing education. Some have called for reconsideration of employing frameworks such as constructivist or experiential learning approaches (Kirschner, Sweller, & Clark, 2006) and others in simulation-based nursing education have called for similar reconsideration (Sanford, 2010; Schiavenato, 2009). Therefore, application of CLT to nursing education research presents a possible and promising avenue for future exploration.

**Implications and Future Directions for Nursing Education Research**

This review provides an overview of Cognitive Load Theory and presents principles that have been explored in cognitive and education psychology. CLT as a framework for instructional design and evaluation is the subject of increasing attention in health professions education research, particularly within the realm of multimedia and simulation-based learning applications. Nurse educators and researchers can contribute to education research by applying principles of cognitive load theory to evaluate instructional design and educational effectiveness.
References


doi:10.1126/science.1736359


doi:10.1111/j.1365-2923.2011.03994.x


Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review, 63*(2), 81.


Chapter 2
Exploration of Context-Based and Non-Context-Based Instruction on Auscultation Skill Development

Background

Health professional students acquire a range of knowledge, skills, and attitudes during their education in preparation for future clinical practice. Developing accurate cardiac and respiratory assessment and auscultation skills form part of the foundational practices of a health professional student’s education. However, both cardiac and respiratory systems have been identified as requiring greater attention given the general weaknesses demonstrated in these skills by health professional students and trainees (Holmboe, 2004; Mangione & Nieman, 1997, 1999; Mangione, 2001).

As health professionals, undergraduate nursing students must also learn these skills in order to practice in a safe and effective manner. Nurses must be able to identify and report to the interprofessional team the physical assessment findings of patients. While the scope of practice for an undergraduate-prepared nurse does not include clinical diagnosis, the ability to identify physical findings is important for patient assessment and management. Because abnormal auscultation sounds are not routinely encountered in actual patients, opportunities for exposure to a wide range of normal and abnormal cardiac and respiratory findings are limited. Therefore, many learning and technology aids have been used to provide this instruction; this includes the use of standardized patients, computer-based animations, and human patient simulators.

The patient simulator has been used with increasing frequency in nursing education (Jeffries, 2005), and human patient simulators (SIMs) have been incorporated into nursing education programs to provide students with clinical skills practice and to
function as surrogates for actual patient encounters (Nehring, 2008). Many studies have advocated the use of SIM in undergraduate nursing education, with high learner satisfaction and increased self-confidence as primary motivators for use (Bambini, Washburn, & Perkins, 2009; Blum, Borglund, & Parcells, 2010; Fountain & Alfred, 2009; Garrett, MacPhee, & Jackson, 2010; Leigh, 2008; Nehring & Lashley, 2009; Reilly & Spratt, 2007; Schoening, Sittner, & Todd, 2006; Weaver, 2011; Wotton, Davis, Button, & Kelton, 2010). Others have argued that clinical nursing performance improves as a result of SIM exposure (Alinier, Hunt, Gordon, & Harwood, 2006; Alinier, Hunt, & Gordon, 2004; Bambini et al., 2009; Blum et al., 2010; Weaver, 2011).

Authentic or context-based learning theory offers a framework for understanding the rationale behind SIM-based applications. These learning opportunities provide students with clinical experiences that are emotionally engaging and increase student motivation for learning (Kaakinen & Arwood, 2009; Paige & Daley, 2009; Solnick & Weiss, 2007; Wilson, Shepherd, Kelly, & Pitzner, 2005). Review articles and meta-analyses encompassing all health professional education programs have suggested that SIM-based instruction increases learner satisfaction and clinical knowledge, skills, and behaviors when compared with no instruction (Cook et al., 2011) and when compared to other instructional formats (Cook et al., 2012). Furthermore, providing instruction in authentic and context-based environments is a reasonable approach given that context is an important component of instruction for performance (Godden & Baddeley, 1975). These studies supporting context-based learning environment provide the theoretical rationale for using technology and other learning aids such as SIM for nursing education.
However, results from other studies suggest that context-based authentic learning environments might not uniformly yield improved learning outcomes. Furthermore, if instructional interventions are evaluated through the framework of Cognitive Load Theory, then the focus on context-based learning environments may potentially have a negative impact on learning due to the increased cognitive load placed on learners during instruction. La Rochelle et al (2011) investigated the effects of varying the level of authenticity of learning contexts for medical students. These students received instruction on the topics of abdominal pain, polyuria, and anemia and the hypothesis was that increasing the authenticity of the learning environment (i.e. increasing context-based learning) would result in improved clinical reasoning. Students found that the higher-authenticity learning environments were more emotionally engaging. However, students randomized between receiving paper-based instruction (lowest in authenticity, but with decreased extraneous cognitive load), DVD-based instruction, or standardized patient-based instruction (highest in authenticity, but with greatest perceived extraneous cognitive load) demonstrated no significant difference in clinical reasoning outcomes.

A brief summary of Cognitive Load Theory (CLT) and of the impact cognitive load has on working memory is helpful to understand the theoretical foundations of La Rochelle et al’s (2011) study arguments. During instruction, multiple sources of information input must be processed within working memory. These inputs can be visual, auditory, or tactile and require cognitive processing within the structures of working memory (Mayer, 2010). However, a learner’s working memory is limited (Baddeley, 2010) and any form of instruction will present a cognitive load burden on the learner’s working memory (Sweller, 1994). In CLT, there are three cognitive load variables that
contribute to this working memory burden: the intrinsic cognitive load, extraneous cognitive load, and the germane cognitive load (Van Merriënboer & Sweller, 2010). Intrinsic cognitive load describes the defined learning goal or task, and includes factors such as element complexity and interactivity as well as the level of learner expertise (Leahy & Sweller, 2004). Extraneous cognitive load results from the features and format of instruction that are not necessary for instruction and which therefore impose a greater cognitive processing burden on a learner’s working memory (Paas, Renkl, & Sweller, 2003). Germane cognitive load describes the “generative cognitive processing” that helps a learner understand the instructional material and which promotes retention in long term memory (Mayer, 2010).

One primary goal of effective instructional design is to decrease extraneous cognitive load by removing features or components of instruction that might distract from the aims of instruction. CLT assumes an additive model for all three variables; cognitive overload and ineffective learning result when the sum of these variables exceeds the capacity of working memory (Van Merriënboer & Sweller, 2010).

**CLT as an alternate framework for evaluating SIM education applications.** Increasing the complexity by situating the instruction ‘in context’ can contribute to extraneous cognitive load if the sensory inputs provided by the instructional tools are not appropriately aligned with the objectives of instruction. If the instruction is simple, there is less concern about extraneous load leading to cognitive overload because there is more capacity in working memory to handle the extraneous and unnecessary input. However, if the instruction is more complex and possesses a greater number of components that the learner has to maintain simultaneously in working memory (i.e. greater element
interactivity), then steps need to be taken to decrease extraneous cognitive load so that
the total cognitive burden does not exceed working memory capacity. The literature has
demonstrated that approaches to decrease extraneous cognitive load has positive effects
on learning and transfer (Sweller & Chandler, 1991; Sweller, Tindall-Ford, & Chandler,
1997; Van Merriënboer & Sweller, 2005).

Therefore, it is possible that high-context-based learning applications to help
novice learners develop skills such as patient physical assessment may contribute to
increased extraneous cognitive load on the learner. Within an authentic, context-based
environment, the learner is required to process multiple sensory stimuli that, if
unnecessary to the learning but which are intended to enhance emotional engagement and
to promote ‘realism’, may obfuscate the primary goals of instruction and may render the
student’s learning less effective. The following study was conducted to explore the
effectiveness of these two approaches within the clinical education and practice domain
of cardiac and respiratory auscultation and physical assessment skill development.

Research Questions

Two primary research questions were addressed in this study: I) Is a Context-
Based (CB+) medium of instruction more effective for learning basic pediatric cardiac
and respiratory auscultation sounds when compared to a Non-Context-Based (CB-)
medium of instruction? II) Does Context-Based (CB+) instruction improve subsequent
SIM-based pediatric physical assessment and auscultation performance when compared
to those receiving Non-Context-Based (CB-) instruction?

The hypothesis was that the CB- instructional format would be superior to the
CB+ format for auscultation test performance due to the decreased extraneous cognitive
load experienced by the CB- instruction group as compared with the CB+ group.

However, it was expected that the CB+ instruction group would demonstrate improved clinical assessment performance over the CB- group on the SIM-based clinical scenarios given that the CB+ group’s prior instruction was situated within a more authentic, albeit simulated, clinical environment.

Methods

See Figure 1 for an overview of the study design.

![Figure 1. Study Design](image)

**Participants.** Senior level undergraduate nursing students (n = 60) were recruited for this study. Participants randomized to the CB+, CB-, and Control groups were recruited in the summer of 2008 and 2009. CB+ and CB- participants received instruction on the following six (6) pediatric Cardiac sounds: normal, systolic murmur (Aortic
Stenosis), diastolic murmur (Aortic Insufficiency), continuous murmur (Patent Ductus Arteriosus), mid-systolic click (Mitral Valve Prolapse), holosystolic murmur (Ventricular Septal Defect). CB+ and CB- participants also received instruction on the following five (5) Respiratory sounds: normal, crackles, rhonchi, wheeze, stridor. The Control group received no instruction.

**Research ethics.** The joint Research Ethics Board of McMaster University Faculty of Health Sciences, Hamilton Health Sciences, and St. Joseph’s Healthcare Hamilton approved this study. All participants provided written and informed consent to participate in this study.

**Context-based (CB+) and non-context-based (CB-) instruction groups and Control group.**

**Context-based instruction (CB+) description**

Those randomized to the CB+ group entered a clinical simulation room that was designed to replicate an actual pediatric patient’s hospital room. The pediatric simulators were lying in bed (the child SIM) and in a crib (the infant SIM) and all SIMs were connected to cardiac monitors, an IV, and were covered in a gown. Participants randomized to the CB+ reviewed a pre-set list of cardiac and respiratory sound examples found on the Laerdal® SimBaby™ and MegaCode Kid™. Both the infant and child SIMs were used to cover all pediatric cardiac and respiratory diagnoses used for instruction. Sounds heard on the Infant simulator were the following: normal heart and breath sounds, systolic murmur, continuous murmur, holosystolic murmur, crackles, stridor. Sounds heard on the Child simulator were the following: normal heart and breath sounds, systolic murmur, diastolic murmur, mid-systolic click, crackles, rhonchi, wheeze.
All participants were required to listen to the sounds in a prescribed order, with cardiac sounds presented first and respiratory sounds second. A research assistant was responsible for changing the sounds for participants. Participants listened to the sounds through a stethoscope placed on the SIM’s chest and participants were allowed to practice any other aspects of cardiac and respiratory assessment during this instruction time. Instruction time was standardized to 40 minutes in total to learn all cardiac and respiratory sounds. Beyond providing the names of the cardiac and respiratory sounds listed above to participants, there was no further description or instruction provided for the sounds.

**Non-context-based instruction (CB-) description**

Those randomized to the CB- group entered a clinical simulation room that contained a laptop computer and a set of headphones. No simulators or other patient equipment was in the room beyond an empty hospital bed and desk. The computer was placed on the desk. Participants randomized to the CB- reviewed a pre-set list of cardiac and respiratory diagnosis examples played through the computer via recorded sound files. The sound files were obtained from a database of cardiac and respiratory sounds accumulated for the purposes of instruction, and are sounds that are available through internet sources. All participants in the CB- group listened to the sounds in a prescribed order, with cardiac sounds presented first. The CB- sounds were different from the simulator-based sounds heard by the CB+ group. A research assistant was available to direct participants on how to begin the instruction period, but otherwise, provided no additional instruction on sounds. Instruction time was standardized to 40 minutes in total to learn all cardiac and respiratory sounds. Beyond providing the names of the cardiac
and respiratory sounds listed above to participants, there was no further description or
instruction provided for the sounds.

Control group instruction

The participants assigned to the Control group did not receive any instruction on
cardiac and respiratory sounds. Control group participants only completed the Context-
based Clinical Assessment and Non-Context-based Auscultation Test.

Primary outcome measures.

Non-context-based outcome (CB-O) measure: auscultation test of sounds

Participants assigned to the CB+ (n = 21), CB- (n = 23), and Control groups (n = 16) completed the auscultation test of sounds, the non-context-based outcome (CB-O) measure. Completion rate of this outcome measure was 100%. For the auscultation test of sounds, the same diagnoses used during instruction were used for testing. The source of the sound examples used for cardiac and respiratory testing were as follows: 6 examples of the sounds learned were taken from the simulators (i.e. the sounds projected through the SIM were recorded, converted into audio files, and included in the auscultation test), 6 examples were taken from sound files presented to the CB- instruction group (i.e. the sounds presented through the computer during instruction), and 8 sounds were examples that were unfamiliar to both the CB+ and CB- groups (i.e. the examples of the sound differed from what either instruction groups heard). There were 20 total Cardiac and Respiratory sounds tested. There were no new diagnoses presented during the auscultation test; participants from the CB+ and CB- had been introduced to all diagnoses during the instruction period.
The cardiac diagnoses tested were as follows: normal, systolic murmur (Aortic Stenosis), diastolic murmur (Aortic Insufficiency), continuous murmur (Patent Ductus Arteriosus), mid-systolic click (Mitral Valve Prolapse), holosystolic murmur (Ventricular Septal Defect). Respiratory diagnoses tested were normal, crackles, rhonchi, wheeze and stridor. In scoring participant responses to each sound presented on the auscultation test, each correct response received one (1) point and an incorrect response received zero (0) points. All scores were calculated and reported as percentage of total correct (0-100%).

**Context-based outcome (CB+O) measures: 1) Clinical assessment performance on simulated clinical scenarios; and 2) simulator-based auscultation test**

Forty-two participants (n = 54; 90%) completed some or all of the context-based outcome measures of this study. Two (2) participants from the CB- group and four (4) participants in the Control group did not complete any CB+O measures. The first context-based outcome measure (CB+O measure 1) was physical assessment performance in simulated clinical scenarios of an infant and a child. The second context-based outcome measure (CB+O measure 2) was another auscultation test of sounds with the Infant and Child SIMs. Therefore, there were two auscultation tests: one within a non-context based environment (CB-O measure, detailed above) and another within a context-based environment and situated within an infant and child clinical scenario (CB+O measure 2).

**CB+O measure 1) Clinical assessment performance on two simulated clinical scenarios**

Aside from the 2 CB- participants and 4 Control group participants that chose not to complete this part of the study, all participants completed the two simulated clinical
scenarios. Participants were introduced to two clinical simulations, one infant and one child, and were asked to perform a cardiac and respiratory assessment on each. A brief description was provided to participants regarding the simulation scenario prior to participants entering the patient room. Participants were given 5 minutes to complete the clinical assessment for each scenario.

Different high-fidelity patient simulators were used for the clinical simulations outcome measure than were used during the instruction period for the CB+ group. This was to control for the possibility of improved performance scores in the CB+ instruction group simply because the CB+ participants had previous exposure and practice time on the simulators used for testing. Therefore, the Infant BabySIM® and Child PediaSIM® produced by METI® were used for the simulation scenarios. The infant was an 8-week-old with signs of congestive heart failure from a moderate Ventricular Septal Defect (VSD). The child was a 6-year old with a right pneumothorax and a systolic murmur following a motor vehicle collision. Participant performances in the simulated scenario were video-recorded and scenarios were reviewed and rated by an outside expert clinician unfamiliar with the participants. The rater was blinded to the instruction group to which the participants were assigned. While construct validity was high with the measurement scale (i.e. the ability to differentiate between expert and novice nurses; p < 0.001), the overall generalizability of the scale was fair to moderate, with a Generalizability Coefficient of 0.40. Inter-rater generalizability was 0.30 and Inter-scenario generalizability was 0.74. For further details on the development and psychometric testing of the assessment measure, please see the appendix.
Participants were assigned a global rating of performance based on a 7-point Likert scale (1=poor, 7=excellent) for each cardiac and respiratory assessment on the infant and child scenarios. The Infant simulation and Child simulation performance scores were each calculated based on the average Cardiac and Respiratory assessment scores assigned by the outside rater of the participant performances.

_**CB+O measure 2) Auscultation test on simulators**_

The auscultation test on the simulators consisted of participants identifying the normal and abnormal cardiac and respiratory sounds on the infant and child SIMs. This auscultation test was similar to the previous auscultation test in the CB-O measure, but the test was now within a context-based setting and sounds were presented on the SIMs rather than on a computer. After completing the physical assessment in each simulated clinical scenario, participants were asked to identify the respiratory and cardiac auscultation sounds heard on the infant and child simulators. The infant possessed right- and left-sided crackles and a holosystolic murmur; the child possessed absent breath sounds in the right lungs, clear breath sounds in the left, and a systolic murmur. Each correct response received one (1) point and an incorrect response received zero (0) points for a total of 6 possible points. Neither group was exposed to these two simulators’ sounds during the instruction period or in the CB-O auscultation test. All scores were calculated and reported as percentage of total correct (0-100%). Only the CB+ and CB-instruction group participants completed the auscultation test on simulators.

_**Statistical analyses.**_ IBM SPSS Statistics version 20.0 was used for all analyses. One-way and repeated measures ANOVA were used to compare performance outcomes between instruction groups and when comparing Cardiac and Respiratory sounds.
performance. Tukey’s HSD was used to compare mean differences between instruction group outcomes. Alpha level was 0.05 and a p-value < 0.05 was considered statistically significant.

Results

Non-Context-Based Outcome (CB-O) measure: auscultation test. One-way ANOVA was performed to compare Total Scores and repeated measures ANOVA was used to compare the Cardiac and Respiratory Subscores on the auscultation test between the CB+ Instruction, CB- Instruction, and the Control groups. See Table 1 for mean (SE) of Total Scores and Cardiac and Respiratory Subscores for the instruction groups. Total Scores between instruction groups was statistically significant with F(2,57) = 21.98; p < 0.001. Tukey’s HSD showed statistically significant mean differences in Total Scores between groups. Comparing Total Scores for CB+ to CB-, mean difference (SE) was 10.93 (3.53), p = 0.008; CB+ to Control was 14.29 (3.88), p = 0.001; CB- to Control was 25.22 (3.81), p < 0.001. See Figure 2. There was also a significant interaction between Instruction Group and Cardiac/Respiratory subscore; F(2,57) = 6.00; p = 0.004.

Table 1. Total Auscultation Test Scores and Cardiac and Respiratory Test Subscores Between Instruction Groups

<table>
<thead>
<tr>
<th></th>
<th>CB+ Instruction Group</th>
<th>CB- Instruction Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Score</td>
<td>46.79 (2.55)</td>
<td>57.72 (2.44)</td>
<td>32.50 (2.92)</td>
</tr>
<tr>
<td>Cardiac Subscore</td>
<td>45.00 (2.70)</td>
<td>48.26 (2.58)</td>
<td>20.00 (3.09)</td>
</tr>
<tr>
<td>Respiratory Subscore</td>
<td>48.57 (3.88)</td>
<td>67.17 (3.71)</td>
<td>45.00 (4.44)</td>
</tr>
</tbody>
</table>
Given the Cardiac/Respiratory Subscore and Instructional Group interaction, separate analyses of Cardiac and Respiratory subscores were conducted. One-way ANOVA for Cardiac subscores comparing the CB+, CB-, and Control groups demonstrated an $F(2,57) = 27.7; p < 0.001$ between groups. Respiratory subscores ANOVA demonstrated an $F(2,57) = 9.3; p < 0.001$ between groups. One-way ANOVA of the Cardiac Subscores between the Control group and CB+ group was $F(1,42)=6.3$; Tukey’s mean difference (SE) was -25.00 (4.11); $p < 0.001$. Cardiac Subscores for the Control group and CB- group was $F(1,42)=7.1$; Tukey’s mean difference (SE) was -28.30 (4.03); $p < 0.001$. One-way ANOVA of the Respiratory Subscores between the Control group and CB+ group was $F(1,42)=0.57$; Tukey’s mean difference (SE) was -3.60 (5.90); $p=0.58$. Respiratory subscores for the Control group and CB- group was $F(1,42)=3.9; -22.20 (5.78); p<0.001$.

Analyses were performed on Auscultation Test subscores for sound examples that were previously heard during instruction (OLD) and for sound examples that were not previously heard during instruction (NEW). These analyses were performed to
understand differences in performance between recalled sounds (OLD) and near transfer performance (NEW). For the OLD sounds analysis, only the CB+ and CB- results were compared, as the control group did not receive any previous instruction. For the NEW sounds analysis, the CB+, CB-, and Control groups’ results were included. All Subscore analyses were conducted for OLD and NEW Total subscores, as well as for Cardiac and Respiratory sounds separately.

For OLD sounds, a one-way ANOVA demonstrated significantly higher scores for the CB- group compared to the CB+ group for Total subscores of OLD sounds previously heard during the instruction period. Analyses of Cardiac subscores for OLD sounds demonstrate significantly higher scores for the CB- group compared to the CB+ group. There was no difference between the CB+ and CB- instruction groups for Respiratory subscores for OLD sounds. ANOVA demonstrated $F(1,42) = 3.04; p = 0.004$ for Total subscores of OLD sounds; $F(1,42) = 5.68; p < 0.001$ for OLD Cardiac subscores; and $F(1,42) = 0.29; p < 0.77$ for OLD Respiratory subscores. See Table 2.

Table 2. Auscultation Test Subscores for OLD Total, Cardiac, and Respiratory Sounds

<table>
<thead>
<tr>
<th></th>
<th>CB+ Instruction Group</th>
<th>CB- Instruction Group</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLD Total Subscore</strong></td>
<td>Mean% (SD)</td>
<td>Mean% (SD)</td>
<td></td>
</tr>
<tr>
<td>Mean% (SD)</td>
<td>43.65 (3.72)</td>
<td>58.70 (3.29)</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>OLD Cardiac Subscore</strong></td>
<td>Mean% (SD)</td>
<td>Mean% (SD)</td>
<td></td>
</tr>
<tr>
<td>Mean% (SD)</td>
<td>31.75 (2.79)</td>
<td>59.42 (3.90)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><strong>OLD Respiratory Subscore</strong></td>
<td>Mean% (SD)</td>
<td>Mean% (SD)</td>
<td></td>
</tr>
<tr>
<td>Mean% (SD)</td>
<td>55.56 (7.03)</td>
<td>57.97 (4.79)</td>
<td>NS</td>
</tr>
</tbody>
</table>

Lastly, analyses were performed on Auscultation Test subscores for sound examples that were not previously heard during instruction (NEW) to evaluate
participants’ near transfer ability to identify new sound examples accurately. For NEW sounds, a one-way ANOVA demonstrated an $F(2,57) = 17.42; p < 0.001$ between instruction groups. See Table 3 for mean (SE) auscultation test subscores for NEW sounds. There was also an interaction between Instruction group and NEW Cardiac/Respiratory subscore: $F(2,57) = 15.97; p < 0.001$.

Table 3. Auscultation Test Subscores for NEW Total, Cardiac, and Respiratory Sounds

<table>
<thead>
<tr>
<th></th>
<th>CB+ Instruction Group</th>
<th>CB- Instruction Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NEW Total Score</strong></td>
<td>Mean% (SD)</td>
<td>48.13 (2.64)</td>
<td>57.30 (3.07)</td>
</tr>
<tr>
<td><strong>NEW Cardiac Subscore</strong></td>
<td>Mean% (SD)</td>
<td>50.68 (3.30)</td>
<td>43.48 (3.27)</td>
</tr>
<tr>
<td><strong>NEW Respiratory Subscore</strong></td>
<td>Mean% (SD)</td>
<td>45.58 (3.77)</td>
<td>71.12 (3.93)</td>
</tr>
</tbody>
</table>

Tukey’s HSD comparisons between instruction groups revealed statistically significant mean differences in NEW sound subscores between groups. For Total Subscores of NEW sounds, the CB- group consistently demonstrated higher scores when compared to the CB+ group [mean difference (SE) = 9.17 (3.90); $p = 0.022$] and the Control group [mean difference (SE) = 24.80 (4.21); $p < 0.001$]. The CB+ group also demonstrated higher scores when compared with the Control group [mean difference (SE) = 15.63 (4.29); $p = 0.002$].

Analyses of Cardiac subscores for NEW sounds demonstrate comparable scores when comparing the CB- group to the CB+ group: ANOVA demonstrated an $F(1,42) = 1.55; \text{mean difference (SE) = } -7.20 (4.33); p = 0.13$. However, both the CB+ and CB-group scores performed significantly better than the Control group for NEW Cardiac
sounds: $F(1,35) = 6.87; \text{mean difference (SE)} = 30.68 (4.76); p < 0.001$ for the CB+ vs. Control group comparison, and $F(1,37) = 5.19; \text{mean difference (SE)} = 23.48 (4.67); p < 0.001$ for the CB- vs. Control group comparison.

There was no difference between the CB+ and the Control group on Respiratory subscores for NEW sounds; $F(1,35) = 0.10; \text{mean difference (SE)} = 0.58 (6.15); p = 0.93$. However, the CB- group performed significantly better than both the CB+ and Control groups: $F(1,42) = 4.67; \text{mean difference (SE)} = 25.54 (5.59); p < 0.001$ for the CB- to CB+ comparison, and $F(1,37) = 4.18; \text{mean difference (SE)} = 26.12; p < 0.001$ for the CB- to Control group comparison.

**Summary of results for the Non-Context-Based Outcome (CB-O) measure:**

**Auscultation test.** Overall results demonstrated better auscultation test performance for the non-context-based (CB-) instruction group over the context-based (CB+) instruction group, which demonstrated better performance than the Control group. There was also a significant interaction between Instruction group and Cardiac/Respiratory subscores. Whereas both the CB+ and CB- performed better than the Control group for Cardiac sounds, the CB- group performed better than both the CB+ and the Control group for Respiratory sounds.

Further analysis of auscultation test performance on OLD sounds revealed that the CB- group demonstrated better performance than the CB+ group for Cardiac sounds but equal performance on Respiratory sounds. Performance on NEW sounds revealed similar patterns as was found for Overall results: the CB- group demonstrated better performance than the CB+ group which also demonstrated superior performance to the Control group on the NEW auscultation test sounds. Furthermore, CB+ and CB- groups both
demonstrated better performance than the Control group for NEW Cardiac sounds, and
the CB- group demonstrated better performance than both the CB+ and Control groups
for NEW Respiratory sounds.

In summary, for the analyses of the CB-O measure of Auscultation Test
performance, the CB- group consistently demonstrated comparable or superior
performance to the CB+ group, and the CB- was consistently superior to the Control
group in all overall scores and subscore analyses.

**CB+O measure 1) simulated clinical scenario performance.** Repeated
measures ANOVA for clinical performance on the Infant and Child simulation scenarios
was not significant between the Instruction groups; F(2,40) = 0.17; p = 0.85. There was
no interaction between Instruction group and Scenario scores. Clinical performance on
the Infant and Child simulation scenarios revealed a range of performances scores
between 1 to 6 on the 7-point likert scale for all groups, with no apparent ceiling or floor
effect or skewness (skew=0.04) in the distribution of performance scores. See Table 4 for
mean (SE) simulation performance scores across Instruction groups.

<table>
<thead>
<tr>
<th>Table 4. Clinical Simulation Performance Scores Between Instruction Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CB+ Instruction Group</strong></td>
</tr>
<tr>
<td><strong>Overall Scenario Score</strong></td>
</tr>
<tr>
<td><strong>Infant Scenario</strong></td>
</tr>
<tr>
<td><strong>Child Scenario</strong></td>
</tr>
</tbody>
</table>

**CB+O measure 2) auscultation test on simulators.** One-way ANOVA for
performance on the context-based outcome measure of Auscultation Test performance on the simulators was not significant between the CB+ and the CB- instruction groups. Mean (SE) auscultation test scores for the CB+ instruction group was 42.86 (5.22); CB- instruction group mean (SE) score was 51.19 (6.56); F(1,40) = -0.99; p = 0.33. While the CB- instruction group demonstrated a higher average score of almost 10 percentage points in comparison to the CB+ group, the results were not statistically significant.

**Summary of results for the context-based outcome (CB+O) measures:**

**clinical performance and simulator-based auscultation test.** There was no significant difference between participant performance scores on both context-based outcome (CB+O) measures between the context-based (CB+) and non-context-based (CB-) instruction groups.

**Discussion.** Overall results for the auscultation test supported the first hypothesis that the non-context-based (CB-) instruction group would demonstrate superior performance on the non-context-based auscultation test (CB-O measure) when compared to the context-based (CB+) instruction group. There was comparable performance on the auscultation test within the context-based environment (CB+O measure). However, the higher average performance scores of the CB- group over the CB+ group suggests better performance, even though this effect was not statistically significant. Therefore, the first hypothesis, that CB- instruction would yield superior performance on auscultation test when compared to those who received CB+ instruction was generally, although inconsistently, supported. No findings suggested CB- instruction was inferior to CB+ instruction. The CB- group consistently demonstrated superior performance to the Control group on the CB-O auscultation test, whereas the CB+ instruction group did not.
The second hypothesis, that the CB+ instruction group would demonstrate superior performance to the CB- group in the clinical assessments, was not supported; performance between the CB+ and CB- instruction groups was comparable. Furthermore, neither group demonstrated significantly different performance than the Control group on this context-based outcome measure. These findings were not expected. Given the strong literature support for high-fidelity simulation in nursing education as contributing to improved clinical performance, the findings from this study do not validate the recommendations in the nursing education literature.

Another way to interpret these findings is that cardiac and respiratory auscultation instruction in a context-based environment did not improve learning for any outcomes measured in this study; context-based instruction did not improve performance over participants who received instruction in a non-context-based environment. In designing the CB- instruction, the intention was to minimize extraneous cognitive load variables that were felt to hinder the primary instructional goal, the auscultation test of sounds. However, by minimizing extraneous cognitive load, the participants would also be prevented from engaging in the type of learning facilitated by context-based instruction (e.g. tactile manipulation of and interaction with SIMs who possessed features of actual patients, learning within an authentic environment that would more closely resemble what a student would encounter in actual practice). Therefore, having the context-based outcome measure – participant performance on two clinical scenarios – was necessary to determine if there were any detrimental effects to clinical performance resulting from learning within the non-context-based instructional environment.
One key application of CLT principles for instructional design is that extraneous cognitive load variables can be decreased in order to prevent cognitive overload in learners. The concern arising from employing authentic, context-based instructional approaches is that these environments can contribute to extraneous cognitive load; furthermore, simplifying instruction and taking students out of the context-based environments could minimize these extraneous cognitive load variables. Therefore, while learners perceive the authentic contexts to be more emotionally engaging (Kaakinen & Arwood, 2009; La Rochelle et al., 2011), the multi-sensory inputs within these instructional environments may actually hinder or distract the student from achieving the intended learning goals.

An alternative framework to context-based learning, therefore, could be to focus on aligning the learning aids and media used for instruction with the primary instructional goal or task, and to minimize aspects of the instruction that may contribute to extraneous cognitive load. Such an approach can facilitate cognitive processing of instructional material (Cook & Triola, 2009) and is in contrast to instructional design that focuses simply on creating highly context-based, or authentic, learning environments. This is especially relevant given the increasing use of learning technologies such as high-fidelity simulations, human patient simulators, and multimedia learning into health professions education. As stated in Rouet, Levonen, and Biardeau (2001, p.1), “[t]here is a subtle shift of attention from what can be done with the technology to what should be done in order to design meaningful instructional applications …. This shift has involved the technological and pedagogical integration between learner cognition, instructional design, and instructional technology, with much of this integration focusing on the role of
working memory in the development of comprehension and performance.” Controlling cognitive load variables in instructional design and evaluating the impact of these variables on learning outcomes is a useful framework for health professions education research.

High-fidelity simulators use in health professions education may not uniformly be appropriate for all instructional goals. Part of the difficulty with high-fidelity simulator use, perhaps, is that there is an implicit assumption that greater technological sophistication of the learning aid translates into more effective instruction; and the closer to reality the instructional environment or instructional tools are, the better. Therefore, even terms such as ‘high-fidelity’ require further clarification and reconceptualization. Several health professions education researchers have called for the reconceptualization of the term ‘fidelity’ when referring to human patient simulators and simulation-based instructional environments (Dieckmann, Gaba, & Rall, 2007; Norman, Dore, & Grierson, 2012). ‘Fidelity’ has most frequently referred to the how closely the physical aspects of instruction are replicated. Therefore, a high-fidelity simulator more closely replicates the physical features of a patient than a low-fidelity simulator. However, the conceptualization of fidelity needs to be expanded to focus not only on the physical features, but also on the psychological or environmental aspects of instruction (Maran & Glavin, 2003).

Several studies have explored this issue of ‘high’ vs. ‘low’ fidelity learning and have also demonstrated that higher-fidelity environments do not necessarily translate into improved performance outcomes for learners. This has been demonstrated with cardiac auscultation skills using the Harvey simulator (De Giovanni, Roberts, & Norman, 2009).
as well as in surgical procedural skills development (Grober et al., 2004; Matsumoto, Hamstra, Radomski, & Cusimano, 2002; Wanzel, Matsumoto, Hamstra, & Anastakis, 2002). This is not a general statement, however, to discourage the use of human patient simulators or high-fidelity simulation for health professions education. In a meta-analysis comparing technology-enhanced simulation with other instructional methods, if the simulation-based instruction was found to have lower extraneous cognitive load relative to the comparison intervention, then higher effect sizes were found; conversely, there were lower effect sizes when the comparison group had more extraneous cognitive load features than the simulation-based instruction groups (Cook et al., 2012). Therefore, seeking to minimize extraneous cognitive load in the instructional design does not preclude educators’ use of technologically sophisticated learning aids.

**Limitations**

There are several limitations in this study. The sound examples were not consistent between context-based and non-context-based instruction groups. Therefore, any difference in performance on the CB+O and CB-O measures could be due to the participants having received different, and possibly poorer quality, sound examples during the instruction period. Furthermore, with no additional instruction being provided to the context-based instruction (CB+) group with respect to clinical assessment on the simulators, it might be expected that no difference were found in Clinical Assessment Performance (the context-based outcome measure of clinical performance) between instruction groups. Additionally, the measurement scale demonstrated only fair to moderate generalizability; therefore, this scale might not have been the best measurement tool to differentiate between participant performances. Finally, the outcome measures
only focused on post-intervention learning and there was no retention or delayed test to evaluate the effect of instruction on long-term learning; interpretation of long-term learning benefit from these instructional interventions must therefore be made with caution.

**Conclusion**

This study evaluates learning outcomes for students receiving instruction within contexts similar to what would be encountered in future practice. Learning within an authentic or context-based environment did not yield improved performance on all the outcomes measured, even though the instructional setting closely mimicked an actual patient context and with high-fidelity simulators functioning as patient surrogates. In general, the non-context-based instruction groups performed either comparable to or, frequently, better than the context-based instruction group on all outcome measures. The results from this study suggest that highly contextualized learning environments may lead to ineffective learning and contribute to cognitive overload by significantly increasing extraneous cognitive load in novice learners. Further exploration is required to understand the effects of context-based learning for all levels of learners, including non-novice learners. This study demonstrates, however, that CLT is a useful framework for instructional design for health professions education, in general, and for this nursing education domain of developing cardiac and respiratory auscultation skills, in particular.
References


Chapter 3

Manipulation of Cognitive Load Variables and Impact on Auscultation Test Performance

Background

In preparation for future clinical practice, students in health professional programs acquire a range of skills throughout their training and education. Cardiac and respiratory auscultation skill development form part of the core competencies a health professional student must acquire for clinical practice. However, health professional students have demonstrated weakness with auscultation skills within both the cardiac and respiratory systems, and educators have identified this area as requiring greater focus in health professions education (Holmboe, 2004; Mangione & Nieman, 1997, 1999; Mangione, 2001). Undergraduate nursing programs comprise one group within the health professions, and nurses are likewise responsible for demonstrating appropriate cardiac and respiratory auscultation skill in order to enact safe and competent patient care. While the scope of practice for a Registered Nurse does not involve clinical diagnosis, nurses must still identify and report physical assessment findings to the interprofessional team. Therefore, with learning goals such as cardiac and respiratory auscultation skill development, educators can tailor instruction to optimize student learning by applying theories from the field of cognitive psychology. In particular, a basic understanding of the role of working memory in learning and applying principles of cognitive load theory (CLT) could help nurse educators design effective instructional material to promote student learning.
Learners receive instruction and information through multiple sensory channels, including the visual and auditory pathways (Mayer, 2002). This information is processed within the cognitive structures of the learner’s working memory (Baddeley, 2010; Burgess & Hitch, 1999) which interacts with, and produces changes to, long-term memory (Mayer, 2002). There are several assumptions about the cognitive architecture of working memory and long-term memory. The first assumption is that working memory is constrained and limited. Therefore, instruction that exceeds the capacities of a learner’s working memory cannot be retained (Van Merrienboer & Ayres, 2005). The second assumption is that the structures of working memory and long-term memory, which is virtually unlimited, can interact. Therefore, information processed within working memory could be retained in the infinite stores of long-term memory, and likewise, information from long-term memory could be retrieved to interact with and facilitate working memory processes (Schnotz & Kürschner, 2007). The third assumption is that the learner’s cognitive load can be adjusted through instructional design in order to facilitate working memory processes (Mousavi, Low, & Sweller, 1995). Because of the limited working memory of learners, educators can benefit from understanding how to apply cognitive load theory to instructional design to enhance student learning. Managing the cognitive load imposed on a learner during instruction can facilitate the success of instruction (Paas, Tuovinen, Tabbers, & Van Gerven, 2003) whereas poorly designed instruction which exceeds the capacity of a learner’s working memory can lead to cognitive overload and ineffective learning (Doolittle, McNeill, Terry, & Scheer, 2005).

CLT assumes an additive model for the different sources of cognitive load placed on the learner. For any particular learning task or goal, the total cognitive load must not
exceed working memory capacity (Sweller, 1994). The cognitive load variables in CLT are the intrinsic load, the extraneous load, and the germane load (Van Merriënboer & Sweller, 2005). Intrinsic cognitive load speaks to the specific learning goal or task and incorporates factors such as the learner’s level of expertise and the inherent difficulty of the learning goal. Element interactivity, a key component to intrinsic cognitive load, describes the number of isolated components a learner needs to hold simultaneously in working memory when processing the information presented (Leahy & Sweller, 2005). Increasing element interactivity of the learning goal increases the intrinsic cognitive load burden on working memory (Paas, Renkl, & Sweller, 2003). Extraneous cognitive load describes the features of instruction that are neither relevant nor necessary for the particular learning goal or task and, therefore, impose an unnecessary burden on working memory (Sweller, Van Merrienboer, & Paas, 1998). Germane cognitive load also contributes additional cognitive burden to working memory processes, but this load is considered useful as it serves to facilitate learners’ understanding and automation of the information received during instruction into long-term memory (Paas & Van Merriënboer, 1994). While the goal of instruction based on CLT is to decrease extraneous cognitive load which is unnecessary for learning, it is also useful to increase germane cognitive load as long as the total cognitive burden placed on learners through the intrinsic load does not already exceed working memory capacity. Therefore, educators applying the principles of the CLT framework can optimize student learning by maintaining the intrinsic cognitive load within the limits of a learner’s working memory, decreasing the ineffective extraneous load of the instructional design, and facilitating information processing through germane cognitive load in learners.
As intrinsic cognitive load is dependent on learner expertise, so also are learner characteristics important in determining whether instructional design contributes to extraneous cognitive load or fosters germane cognitive load. For example, instructional design that is appropriate for the novice learner and promotes germane cognitive load may be found to contribute only extraneous load in a more expert learner (Leahy & Sweller, 2005). Therefore, concrete and ecologically valid learning contexts, evaluating specific levels of learners, are needed to understand how cognitive load variables interact (Van Merriënboer, Kester, & Paas, 2006).

Cognitive load theory has received increasing attention in the health professions literature and has been promoted as a framework to guide health professions education research (Mayer, 2010; Van Merriënboer & Sweller, 2010). Currently, there are no known studies in nursing education that evaluate the application of CLT to instructional design. The following studies explore the effects of manipulating cognitive load variables of instructional design and evaluate senior-level undergraduate nursing student performance on cardiac and respiratory auscultation tests. While the nursing students in this study are in their final semesters of their health professional program, they are still considered novice learners with respect to the auscultation skill instruction they have received in these studies.

The auscultation tests used in the following studies will evaluate performance outcomes through several measures. While overall auscultation test performance will be the primary outcome measure to assess learning of cardiac and respiratory diagnoses, analysis will also include evaluation of recall and near transfer performance, both of which are subgroups within the overall test performance scores. It is helpful to
understand how, or if, instructional design modifications impact learner performance not only for overall auscultation test outcomes, but whether there is an effect on learners’ ability to transfer learning to other contexts (near or far transfer) and in clinical situations different from the instructional environment. Furthermore, delayed testing are incorporated into these studies given that the goal of instruction is to promote retention of learning following the initial instruction period.

**Study 3.1: Modification of Intrinsic Cognitive Load (ICL) and impact on auscultation test performance**

When designing instructional session for learners, an educator must decide what content will be covered during a learning session. In deciding what content to cover, there is an implicit determination concerning the optimal intrinsic cognitive load that will allow the learners attain the desired learning goals. Because working memory is limited, an increase in the quantity or complexity of instructional content results in a commensurate decrease in the amount of information a learner is able to retain. However, significantly decreasing the instructional content also has drawbacks. While a learner may be able to retain a greater proportion of the content covered during the session, this might be an inefficient use of instructional time given that the total content to be covered must be then delivered in subsequent learning sessions. Therefore, evaluating the effects of modulating the intrinsic cognitive load variable by modifying the units of instruction is worth exploring.

One approach to modifying intrinsic cognitive load is to reduce element interactivity within the instructional task (Pollock, Chandler, & Sweller, 2002). This has shown to be effective in math instruction (Ayres, 2006) and studies demonstrate that
reducing the number of elements that a learner must retain simultaneously in working memory is beneficial in improving overall performance (Bannert, 2002). This allows educators to add complexity to the learning task at later stages of learning. However, there is an issue of ecological validity: as element interactivity continues to be decreased for a complex task, and the instruction sessions cover less and less content, there will be a point where the number of sessions required to cover all the desired content is not actually feasible to implement. The impracticality of teaching only one diagnosis at a time precludes its implementation in health professional education, and educators must also maximize the time available to cover the required material. Studies in other disciplines have also found that focused instruction with minimal element interactivity may yield high performance on subsequent testing, but only when those exact items are being tested. The ability of learners to transfer learning to new contexts is limited when element interactivity is low (Simons, 1999; van Merrienboer, De Croock, & Jelsma, 1997).

The decisions regarding how much content to cover in an instructional session is difficult with complex tasks like auscultation skill development. Educators must evaluate the effects of reducing element interactivity with the need to cover required information necessary for students to have learned the fundamental concepts within a particular learning domain. Therefore, reducing number of diagnoses learned would likely improve performance on those fewer diagnoses. Those results would be expected. However, would reducing element interactivity facilitate the desired goal of near transfer when compared with instructional sessions that have greater element interactivity?
Because of the uncertainty around cardiac and respiratory auscultation skill development and the effects of instruction on subsequent auscultation test performance, the first study was conducted. The purpose of this study was to evaluate performance outcomes by manipulating the intrinsic cognitive load of the learning goal. Furthermore, the goal was to evaluate whether there were additional benefits of minimizing intrinsic cognitive load that resulted in enhanced performance outcomes for near transfer and on delayed testing.

**Research Question**

How does intrinsic cognitive load impact on cardiac and respiratory auscultation test performance when learners are provided with instruction on either three or six diagnoses?

**Methods**

**Participants.** Senior level undergraduate nursing students (n = 44) were recruited for this study. Participants were provided with either Cardiac (n = 21) or Respiratory (n = 23) sounds instruction. The Respiratory sounds instruction group was recruited in early-July, 2010, and the Cardiac sounds instruction group was recruited in late-July, 2010. After participants were identified to receive either Respiratory or Cardiac instruction, they were randomized to receive instruction on either three (“3 Group”) or six (“6 Group”) diagnoses.

**Study design.** Six diagnoses were used for the Respiratory and Cardiac sounds interventions. The following diagnoses were used for Respiratory instruction: Normal Bronchial, Normal Vesicular, Stridor, Wheeze, Crackles, Rhonchi. The following diagnoses were used for Cardiac instruction: Normal, Atrial Septal Defect, Aortic Stenosis, Aortic Insufficiency, Mitral Valve Prolapse, Ventricular Septal Defect.
A group of clinical experts was consulted to determine which three diagnoses from each body system they believed were most difficult to learn. Based on the feedback of the clinical experts, three Cardiac and three Respiratory diagnoses were identified as most difficult of the six. The 3 Group received instruction on these three diagnoses. The 6 Group learned all six diagnoses. The rationale for selecting the three most difficult Cardiac or Respiratory diagnoses was to minimize superior performance by the 3 Group participants simply because they had easier diagnoses to learn.

Time for learning each diagnosis was standardized across both the 3 and 6 instruction groups. For each diagnosis learned, each participant was provided with two minutes of auditory instruction (i.e. listening to the cardiac or respiratory sound) followed by a 10 second rest period before proceeding to the next diagnosis. During the two-minute instruction period, participants were exposed to three different examples of the diagnosis. These diagnosis examples were recorded audio files (i.e. mp3) from a database of cardiac and respiratory sounds, obtained from the Internet through non-proprietary sources. Brief written descriptions accompanied the auditory examples of all sounds provided; the written descriptions were identical for the 3 and 6 Groups.

**Auscultation test.** Following the instruction period, participants were tested with four different examples of each of the diagnoses learned. Two examples were taken from the sounds presented during the instruction period and evaluated participant recall ability (OLD sounds); two examples were not previously heard and evaluated participant near transfer ability (NEW sounds). Therefore, the participants in the 3 Group were tested on twelve total sounds and the participants in the 6 Group were tested on twenty-four sounds. All sound examples were shuffled and the audio files were burned onto a Test
CD. Auscultation test performance scores were calculated based on percentage of total sounds identified correctly. The Respiratory groups received a post-instruction test and a retest of the same sounds one week following instruction. All participants from the Respiratory instruction groups completed the post-instruction test and retest. All Cardiac group participants completed the post-instruction test only.

**Ethics.** The joint Research Ethics Board of McMaster University Faculty of Health Sciences, Hamilton Health Sciences, and St. Joseph’s Healthcare Hamilton approved this study. All participants provided written and informed consent to participate in this study.

**Statistical analysis.** IBM SPSS Statistics version 21.0 was used for the following data analyses. Repeated measures ANOVA was used to compare post-instruction (PI) Cardiac and Respiratory 3 and 6 Group auscultation test performance, subgroup analyses on OLD and NEW sounds performance, and Respiratory test and retest performance. Alpha level was set at $\alpha = 0.05$ with statistical significance considered for p-values < 0.05. Performance outcomes on the three diagnoses common both to the 3 Group and the 6 Group were used in the following analyses. Focusing only on the three diagnoses learned by both groups allows direct comparison of the effect of increasing intrinsic load on auscultation test performance.

**Results**

Overall means (SE) for Auscultation Test performance of the 3 and 6 Groups are found in Table 1.
Table 1. Post – Instruction Auscultation Test Scores, Cardiac and Respiratory Test Subscores, and Respiratory Retest Scores Between Instruction Groups

<table>
<thead>
<tr>
<th>Test</th>
<th>Total Score</th>
<th>Cardiac Subscore</th>
<th>Respiratory Subscore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SE)</td>
<td>Mean (SE)</td>
<td>Mean (SE)</td>
<td>Mean (SE)</td>
</tr>
<tr>
<td>Test</td>
<td>Test</td>
<td>Test</td>
<td>Retest</td>
</tr>
<tr>
<td>“3” Instruction Group</td>
<td>73.48 (3.19)</td>
<td>75.76 (4.64)</td>
<td>71.21 (3.86)</td>
</tr>
<tr>
<td>“6” Instruction Group</td>
<td>57.29 (2.18)</td>
<td>58.33 (4.87)</td>
<td>55.83 (4.05)</td>
</tr>
</tbody>
</table>

**Post-Instruction auscultation test performance.** Repeated measures ANOVA demonstrated a significant difference between the 3 and 6 instruction groups with F(1,21) = 18.06; p < 0.001 for performance on the three diagnoses common to both instruction groups. There was no interaction between the Cardiac or Respiratory instruction groups with the 3 or 6 Groups: F(1,19) = 0.047; p = 0.83; there was also no interaction between OLD/NEW sounds and the 3 or 6 instruction group: F(1,19) = 0.33; p = 0.58. A post hoc analysis using one-way ANOVA was performed to confirm that the three Cardiac and Respiratory diagnoses chosen for the 3 Group were indeed more difficult than the three diagnoses not introduced to the 3 Group. Data from the participants in the 6 Group demonstrated lower average scores on the 3 Group diagnoses for both Cardiac and Respiratory sounds when compared with the remaining three diagnoses not introduced to the 3 Group; the mean difference (SE) between the more and less difficult diagnoses was 7.6 (3.5) percentage points (p = 0.04), with an average Total Score for the remaining three diagnoses of 64.90 for the diagnoses not used in the above analyses.
**Test-Retest auscultation test performance.** Repeated measures ANOVA demonstrated a significant difference between the Respiratory 3 and 6 instruction groups’ performance with the 3 Group achieving higher scores than the 6 Group $F(1,19) = 26.35; p < 0.001$. There was a significant decline in test performance between Test and Retest: $F(1,19) = 31.02; p < 0.001$, and a significant three-way interaction between 3-6 Group, Old-New sounds, and Test-Retest performance: $F(1,19) = 23.74; p < 0.001$.

Data were analyzed further to understand the three-way interaction, and separate analyses of the OLD sounds and NEW sounds were performed. Repeated measures ANOVA of the OLD Respiratory sounds on Test-Retest demonstrated an $F(1,19) = 12.01; p = 0.002$, a significant decline in scores from Test to Retest with $F(1,19) = 52.54; p < 0.001$, and a significant interaction between 3-6 Group and Test-Retest performance with $F(1,19) = 34.73; p < 0.001$. Repeated measures ANOVA of NEW Respiratory sounds between the 3-6 Groups demonstrated an $F(1,19) = 19.34; p < 0001$ with no significant decline between Test and Retest – $F(1,19) = 0.66; p = 0.43$ and no 3-6 Group and Test-Retest interaction – $F(1,19) = 0.29; p = 0.60$. The three-way interaction above, therefore, was carried predominantly by the OLD Respiratory sounds performance on Retest by the 6 Group, which dramatically fell from Test to Retest. See Figure 1.
Summary of results. For the three diagnoses common to both the 3 Group and the 6 Group, the 3 Group consistently demonstrated superior auscultation test performance on the post-instruction test. When evaluating Respiratory auscultation test performance between test and retest, the 3 Group continued to demonstrate superior performance than the 6 Group. Of note, however, whereas the 6 Group performed significantly worse in recalling OLD sounds on retest, there was no significant decline in test performance on NEW sounds between test and retest.

Discussion

This study explored the effect of adjusting intrinsic cognitive load for learning cardiac and respiratory auscultation sounds in novice learners. As expected, those students with fewer diagnoses to learn demonstrated better performance on auscultation test performance than students with twice as many diagnoses to learn. Therefore, increasing intrinsic load by simply increasing the quantity of material to be learned yielded the expected outcomes of decreased performance. However, interestingly, when
looking at performance on NEW sounds between Test and Retest, there was no
significant attenuation in near transfer performance for both the 3 Group and the 6 Group.
These results suggest that, even as the 6 Group learned more diagnoses, their rate of
decline on Retest for near transfer performance was not significantly different from the
rate of decline in the 3 Group. While the 6 Group demonstrated a significant decline on
Retest performance of OLD sounds, this finding is less important given that the more
important goal of instruction is to facilitate transfer, rather than recall, of learning into
new contexts.

In the area of cardiac and respiratory auscultation skill development, other
instructional strategies would be helpful to explore that would evaluate not only changes
to intrinsic cognitive load, but would also address extraneous and germane load variables.
Pragmatically speaking, if an instructional session were able to include twice as much
content (e.g. teaching 6 diagnoses vs. 3 diagnoses), this would be a more efficient use of
instructional time for educators and students alike. However, as the results of this study
demonstrated, the effect of diminished performance must be balanced with the addition
of a greater amount of content. In Ayers’ (2006) study, the focus was on decreasing the
element interactivity of the math calculations to modify intrinsic load math. In this
current study, intrinsic load was modified through varying the number of diagnoses
learned during the instruction period. However, these studies did not explore strategies to
facilitate germane load with learners to facilitate achievement of the learning goals.
Investigating strategies that address the other components of cognitive load would be
worthwhile. Other approaches to modifying cognitive load – in particular, to decrease
extraneous load and facilitate germane load – to optimize delivery of instructional content
and learner performance would allow for a focus on increasing learning and transfer when intrinsic load remains constant. Manipulating extraneous and germane cognitive load variables will be the focus of the next two studies.

**Study 3.2. Contributing to Extraneous cognitive load (ECL) or Fostering germane cognitive load (GCL) through providing multiple representations with instruction.**

Once an instructional goal has been determined, CLT also supports removing aspects of the instruction that are unnecessary for learning – factors which contribute to extraneous cognitive load. Decreasing extraneous load is less important for learning tasks that do not place a high intrinsic load on a learner’s working memory because there are additional working memory resources available to help the learner process both the necessary intrinsic load and the unhelpful extraneous load (Van Merriënboer & Sweller, 2005). However, if the intrinsic cognitive load of a learning task increases, the need to decrease extraneous load is important so that the total cognitive load does not exceed the learner’s working memory capacity. The goal in instructional design is to use the remaining cognitive resources available in working memory by decreasing extraneous load and by facilitating germane load for the particular instructional goal.

Just as there are no standard metrics to quantify the intrinsic cognitive load of an instructional task that does not require consideration of the expertise level of the learner, so too, what constitutes germane load for one level of learner may be unnecessary extraneous load for another (McKeough, Lupart, & Marini, 1995). The definition of germane load also alludes to this difficulty in discerning whether instructional design variables might contribute to extraneous load or facilitate germane load. Education and cognitive psychology researchers acknowledge that germane load is the component of
One such example in the literature is the use of multiple examples, or representations, in instructional design. Adding variability by exposing learners to multiple examples helps learners distinguish relevant from irrelevant features (Van Merriënboer & Sweller, 2005) and facilitates transfer to new contexts (McKeough et al., 1995). One of the first instances reported in the literature around the use of multiple examples to facilitate learning was with instruction in solving algebraic equations of work problems (Reed & Bolstad, 1991). The use of multiple examples, or representations, for instruction has also been found to be beneficial in health professions education with instruction of cardiac physiology principles (Norman, Dore, Krebs, & Neville, 2007; Norman, 2009).

However, educators must exercise caution when using multiple examples in instruction because requiring learners to process multiple representations may increase extraneous cognitive load and be detrimental to learning (Sankey & Nooriafshar, 2005). If a learner does not have experience with the subject matter and has not understood the foundational principles upon which the examples are based, including multiple examples during initial instruction may instead prove detrimental to learning. Therefore, for the following study of instructional design for cardiac and respiratory auscultation skill development, it would be helpful to understand whether the same benefits of providing learners with multiple examples persists. In the first study which compared learning three or six diagnoses, all learners were provided with three examples of each diagnosis.
whether they were randomized to the 3 Group or the 6 Group. The following study explores whether providing only a single example to this group of novice learners might result in decreased performance, especially on near transfer, thereby suggesting that multiple examples were beneficial for identifying cardiac and respiratory auscultation diagnoses. However, if learners were provided with only a single example of each diagnosis and their performance was superior to the group provided with multiple examples, these results might suggest that the multiple examples were adding extraneous cognitive load on learners rather than facilitating germane load.

Research Question

How will providing multiple examples of cardiac and respiratory sounds impact on auscultation test performance over providing only single examples of sounds?

Methods

Participants. Senior level undergraduate nursing students (n = 32) were recruited for this study in June – July, 2011. Participants were randomized to receive instruction with a single example (Single group) or with multiple examples (Multiple group) of the Cardiac and Respiratory diagnoses.

Study design. Eight total diagnoses were used for the Single and Multiple sounds interventions. The following Cardiac diagnoses were used for instruction: Normal, Aortic Stenosis, Aortic Insufficiency, Mitral Valve Prolapse. The following Respiratory diagnoses were used for instruction: Normal Bronchial, Stridor, Wheeze, Crackles. The sound examples provided to the Single group consisted of recorded audio files (i.e. mp3) from the pediatric Laerdal® human patient simulators. These audio files were compiled into an instruction CD for the Single group. The sound examples provided to the Multiple
group consisted of three audio examples of each diagnosis. One example was taken from Laerdal® simulators, identical to what the Single group heard. The two additional examples were taken from a database of cardiac and respiratory sounds. Brief written descriptions accompanied the auditory examples of all sounds provided; the written descriptions were identical for the Single and Multiple groups.

Time for learning each diagnosis was standardized across both the Single and Multiple instruction groups. For each diagnosis learned, each participant was provided with two minutes of auditory instruction (i.e. listening to the cardiac or respiratory sound) followed by a 10 second rest period before proceeding to the next diagnosis. During the two-minute instruction period, participants were exposed to either one example of the diagnosis (Single group) or three different examples of the diagnosis (Multiple group).

**Auscultation test.** Following the instruction period, participants were tested with four different examples of each of the diagnoses learned. For the Single group, the example heard during instruction was included to test for recall (OLD sound) and three additional examples which were not previously heard tested for near transfer (NEW sound). For the Multiple group, two examples were taken from the sounds presented during the instruction period and evaluated participant recall (OLD sounds); two examples were not previously heard and evaluated participant near transfer (NEW sounds). All sound examples used for the auscultation test were shuffled and the audio files were burned onto a Test CD. Auscultation test performance scores were calculated based on percentage of total sounds identified correctly. Participants completed both the post-instruction test and a retest of the same sounds one week following instruction.
Ethics. The joint Research Ethics Board of McMaster University Faculty of Health Sciences, Hamilton Health Sciences, and St. Joseph’s Healthcare Hamilton approved this study. All participants provided written and informed consent to participate in this study.

Statistical analyses and power calculation. IBM SPSS Statistics version 21.0 was used for the following data analyses. Repeated measures ANOVA was used to compare Single and Multiple instruction groups on Cardiac and Respiratory auscultation test performance, for subgroup analyses on OLD and NEW sounds performance, and for test and retest performance. Alpha level was set at $\alpha = 0.05$ with statistical significance considered for p-values < 0.05. Power calculations were performed which demonstrated that for a sample size of 15 participants per instruction group, a difference of 15 percentage points in test performance between intervention groups could be detected with 80% power.

Results

Overall means (SE) Auscultation Test performance scores for the Single and Multiple instruction groups are reported in Table 2.
Table 2. Auscultation Test Scores, Cardiac and Respiratory Test Subscores, and Respiratory Retest Scores Between Instruction Groups

<table>
<thead>
<tr>
<th></th>
<th><strong>Total Score</strong></th>
<th></th>
<th><strong>Cardiac Subscore</strong></th>
<th></th>
<th><strong>Respiratory Subscore</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Mean (SE)</strong></td>
<td><strong>Mean (SE)</strong></td>
<td><strong>Mean (SE)</strong></td>
<td><strong>Mean (SE)</strong></td>
<td><strong>Mean (SE)</strong></td>
</tr>
<tr>
<td><strong>Single Instruction Group</strong></td>
<td>62.55 (2.69)</td>
<td>60.82 (3.73)</td>
<td>61.77 (3.54)</td>
<td>62.50 (5.33)</td>
<td>62.61 (3.25)</td>
</tr>
<tr>
<td></td>
<td><strong>Test</strong></td>
<td><strong>Retest</strong></td>
<td><strong>Test</strong></td>
<td><strong>Retest</strong></td>
<td><strong>Test</strong></td>
</tr>
<tr>
<td><strong>Multiple Instruction Group</strong></td>
<td>59.81 (5.86)</td>
<td>63.33 (5.11)</td>
<td>59.14 (7.41)</td>
<td>60.58 (6.69)</td>
<td>60.48 (5.70)</td>
</tr>
<tr>
<td></td>
<td><strong>Test</strong></td>
<td><strong>Retest</strong></td>
<td><strong>Test</strong></td>
<td><strong>Retest</strong></td>
<td><strong>Test</strong></td>
</tr>
</tbody>
</table>

Repeated measures ANOVA demonstrated a non-significant difference between Single and Multiple group scores: F(1, 28) = 0.000; p = 0.99. Furthermore, there was no significant decline in scores from Test to Retest: F(1, 28) = 0.13; p = 0.73; there was also no interaction between the Single or Multiple instruction group and Test/Retest scores. No further interactions were found between instruction group with Cardiac or Respiratory performance as well as with recall (OLD) or near transfer (NEW) performance.

**Discussion**

These results suggest that for novice nursing students, multiple representations of cardiac and respiratory auscultation diagnoses may either contribute to extraneous cognitive load or may have no additional instructional benefit over providing students with single examples of the diagnoses. It is important to note that students were not provided with any additional instruction around the diagnoses presented. Therefore, there was no instruction regarding how these examples highlighted the defining features of each cardiac or respiratory diagnosis. These results underscore, perhaps, the need for
further explanatory information to accompany the basic written instructions when multiple auditory examples are provided in auscultation instruction. The non-significant findings in this study, however, could also be due to a type-II error due to an insufficient sample size. Therefore, results from this study must be interpreted with caution. While there is sufficient justification in the literature around the use of multiple representations with instruction, the findings were not replicated in this learning context and with this level of learner. The next study investigates another approach that has been discussed in the literature that has been found to facilitate germane load: interleaving examples during instruction.

**Study 3.3. Fostering germane cognitive load (GCL) through an interleaving (or, mixed format) instructional approach.**

The interleaving approach for instruction has been well documented in the cognitive psychology literature, and research on this approach spans several decades. The definition of interleaving instruction, and the focus of this next study, is when two or more learning goals representing a shared learning domain are alternated during instruction (Richland, Bjork, Finley, & Linn, 2005). In the field of auscultation skills development, interleaving would involve alternating examples from two or more diagnoses during instruction. Diagnoses that were not alternated, and examples from each diagnosis presented in succession before moving on to the next diagnosis, would be an example of blocked instruction. Therefore, in the following study, the interleaving approach will be termed “mixed instruction” and the non-interleaving approach will be termed “blocked instruction”.

Interleaving approaches in education were first documented by Shea and Morgan over two decades ago (Shea & Morgan, 1979) for motor skills acquisition. Further studies in motor learning followed and the benefits of the interleaving approach, alternatively termed “mixed practice” or “practice involving contextual interference”, has been demonstrated for a variety of motor skills instruction, from children throwing beanbags (Carson & Wiegand, 1979), to baseball players’ batting skills (Hall, Domingues, & Cavazos, 1994), to basketball players’ basket-shooting skills (Landin & Hebert, 1997). In all the above examples from the motor learning literature, the beneficial effects of interleaving have been demonstrated. While blocked instruction may demonstrate better outcomes during the practice session over mixed instruction (Bjork, 1994, 1999; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010), the benefits of mixed instruction are revealed over time and on delayed testing.

In addition to the initial appearances of blocked instruction being superior to mixed instruction during the practice and immediate post-instruction testing period, learners also perceived blocked instruction to be more effective for their learning than mixed instruction (Birnbaum et al., 2012; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010; Simon & Bjork, 2001). Therefore, even as learners believed that the blocked instruction format was more effective for their own learning and the mixed instruction group perceiving the instructional format to be less effective, their performance results demonstrated the opposite.

In cognitive psychology, the interleaving instructional approach has been termed a “desirable difficulty” in learning (Bjork, 1994), where cognitive load is increased for the learner, but in a manner that facilitates, rather than hinders, learning. Inducing these
“desirable difficulties” in learning facilitate germane cognitive load even as instruction is more taxing to working memory when compared with a blocked instruction approach (Rohrer & Pashler, 2010). The benefits of the interleaving approach have been found with instruction around differentiating painting styles (Kang & Pashler, 2012; Kornell & Bjork, 2008), learning bird and butterfly species (Birnbaum, Kornell, Bjork, & Bjork, 2012), understanding astronomy concepts (Richland et al., 2005), and applying mathematical concepts (Mayfield & Chase, 2002; Rohrer & Taylor, 2007; Taylor & Rohrer, 2010). In the health professions education literature, interleaving or mixed practice during instruction has been helpful in the instruction of psychiatric diagnoses (Zulkiply, McLean, Burt, & Bath, 2012) and in electrocardiogram interpretation (Hatala, Brooks, & Norman, 2003). Clearly, there is much literature around the positive effects of interleaving approaches during instruction.

While this approach has been documented with motor and cognitive skills acquisition, questions remain whether these approaches would also be beneficial for learners receiving auditory instruction for goals such as cardiac and respiratory auscultation skill development. The theoretical rationale for the benefits of interleaving for cognitive goals is that this approach encourages learners to make contrasts between the examples of the learning goals provided. The following study explored whether the interleaving approach for auscultation skill development would, similarly, allow novice health professional leaners to make auditory contrasts between examples to enhance learning of cardiac and respiratory diagnoses.

**Research Question**
How does providing examples of cardiac and respiratory sounds in a mixed format impact on auscultation test performance over providing blocked examples of sounds?

**Methods**

**Participants.** Senior level undergraduate nursing students (n = 22) were recruited for this study in May – June, 2012. Participants were randomized to receive instruction with diagnoses examples presented in an interleaved format (Mixed group) or with diagnoses presented together in succession (Blocked group) of the Cardiac and Respiratory diagnoses.

**Study design.** Eight total diagnoses were used for the Mixed and Blocked group interventions. The following Cardiac diagnoses were used for instruction: Normal, Aortic Stenosis, Aortic Insufficiency, Mitral Valve Prolapse. The following Respiratory diagnoses were used for instruction: Normal Bronchial, Stridor, Wheeze, Crackles. For the Blocked instruction group, three examples of each diagnosis were provided in succession. All three examples of one diagnosis were played before moving to next diagnosis (i.e. Diagnosis 1, examples A, B, C; Diagnosis 2, examples A, B, C; etc). For the Mixed instruction group, three examples of each diagnosis were provided. However the diagnoses were presented in pairs and the examples of each diagnosis were interleaved (i.e. Diagnosis 1, example A; Diagnosis 2, example A; Diagnosis 1, example B, C; Diagnosis 2, example B, C; Diagnosis 3, example A; Diagnosis 4, example A; Diagnosis 3, example B, C; Diagnosis 4, example B, C; etc).

The sound examples were audio files taken from Laerdal® simulators and from a database of cardiac and respiratory sounds. All examples for each diagnosis were the same for the Mixed and Blocked instruction group; only the order in which the examples
were presented was changed. Brief written descriptions accompanied the auditory examples of all sounds provided; the written descriptions were identical for the Mixed and Blocked instruction groups. Time for learning each diagnosis was standardized across both the Mixed and Blocked instruction groups. For each diagnosis learned, each participant was provided with two minutes of total auditory instruction (i.e. listening to the cardiac or respiratory sound).

**Auscultation test.** Following the instruction period, participants were tested with four different examples of each of the diagnoses learned. Two Auscultation Test sounds for each diagnosis were taken from the examples presented during the instruction period and evaluated participant recall (OLD sounds); two examples were not previously heard and evaluated participant near transfer (NEW sounds). All sound examples used for the auscultation test were shuffled and the audio files were burned onto a Test CD. Auscultation test performance scores were calculated based on percentage of total sounds identified correctly. Participants completed both the post-instruction test and a retest of the same sounds one week following instruction.

**Research ethics.** The joint Research Ethics Board of McMaster University Faculty of Health Sciences, Hamilton Health Sciences, and St. Joseph’s Healthcare Hamilton approved this study. All participants provided written and informed consent to participate in this study.

**Statistical analyses.** IBM SPSS Statistics version 21.0 was used for the following data analyses. Repeated measures ANOVA was used to compare Mixed and Blocked instruction groups on Cardiac and Respiratory auscultation test performance, for subgroup analyses on OLD and NEW sounds performance, and for test and retest
performance. Alpha level was set at $\alpha = 0.05$ with statistical significance considered for p-values < 0.05.

**Results**

Overall means (SE) Auscultation Test performance scores for the Mixed and Blocked instruction groups are reported in Table 3.

Table 3. Auscultation Test Scores, Cardiac and Respiratory Test Subscores, and Respiratory Retest Scores Between Instruction Groups

<table>
<thead>
<tr>
<th></th>
<th>Total Score Mean (SE)</th>
<th>Cardiac Subscore Mean (SE)</th>
<th>Respiratory Subscore Mean (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test</td>
<td>Retest</td>
<td>Test</td>
</tr>
<tr>
<td><strong>Blocked Instruction Group</strong></td>
<td>49.91 (4.30)</td>
<td>49.18 (4.48)</td>
<td>44.27 (4.71)</td>
</tr>
<tr>
<td><strong>Mixed Instruction Group</strong></td>
<td>71.06 (3.25)</td>
<td>64.17 (2.70)</td>
<td>64.38 (3.85)</td>
</tr>
</tbody>
</table>

Repeated measures ANOVA demonstrated a significant difference in Auscultation Test performance between the Mixed and Blocked instruction groups, with the Mixed instruction group consistently performing significantly better than the Blocked instruction group on Total score and for the Cardiac and Respiratory subscores: $F(1,20) = 15.95; p = 0.001$. There was no interaction between Mixed or Blocked instruction on Test and Retest: $F(1,20) = 0.01; p = 0.93$; while all scores decreased on Retest, the decrease in scores was not significant: $F(1,20) = 3.67; p = 0.07$.

However, a significant finding was that there was an interaction between Mixed or Blocked instruction and recall (OLD) and near transfer (NEW) performance: $F(1,20) =$
Discussion

Results from this study for cardiac and respiratory auscultation test performance are consistent with the literature around the beneficial learning effects of interleaving. Furthermore, the interleaving approach not only improved overall test performance, but the interaction between OLD and NEW sounds further suggests that interleaving may have less effect on learners’ recall abilities, but helped to facilitate near transfer. If the goal of education is to help learners not simply remember information or knowledge that has been presented, but to apply the learning to new examples or contexts, the results from this study suggest that interleaving facilitates such near transfer. These findings regarding interleaving for near transfer support the previously published research of instruction on painting styles differentiation (Kang & Pashler, 2012). The findings in this study suggest that for the novice learner of cardiac and respiratory diagnoses, mixed instruction through interleaving supported learning by facilitating germane cognitive load.
over blocked instruction. Therefore, to contrast this study with the previous exploration around Single or Multiple examples, these finding suggest that multiple examples, without further guidance or elaboration, may contribute to extraneous cognitive load while mixed instruction promoted germane cognitive processing by allowing learners to make active comparisons between the examples, even when no further instruction was provided.

Education researchers have speculated that the interleaving approach encourages learners to differentiate and discriminate between examples (Kang & Pashler, 2012; Taylor & Rohrer, 2010). The act of juxtaposing examples from different categories (e.g. wheezes vs. crackles) within a learning domain (e.g. respiratory diagnoses) encourages learners to recognize differences and enhances inductive learning (Birnbaum et al., 2012). The discriminative contrast hypothesis postulates that interleaving promotes inductive learning whereas temporal spacing does not (Kang & Pashler, 2012; Kornell & Bjork, 2008). Throughout this study, the definition of interleaving was intentionally specific regarding the alternating of examples within a shared learning domain. The reason for this specificity was because the instructional approach could be easily confounded with a similar instructional approach, but one not explored in this study, of temporal spacing whereby examples from a non-shared domain were contrasted (e.g. providing an example of wheeze, then an auditory example of a classical music piece, then a different example of a wheeze). In addition to the two studies mentioned above, the effects of spacing and distributed practice have been explored elsewhere and have demonstrated variable learning outcomes (Birnbaum et al., 2012; Richland et al., 2005; Rohrer & Pashler, 2010; Rohrer & Taylor, 2007); these effects, however, have not been investigated in this study.
The findings in this study have implications for auscultation skill development for health professional students. Mixed, or interleaved, instruction may introduce increased cognitive load on a learner’s working memory, but this increased cognitive load was found to contribute to germane cognitive processing and facilitated near transfer performance. Blocked instruction may be perceived by learners to be more effective, but this perception has not been supported by the literature as contributing to better learning outcomes. It is likely that the positive perceptions around blocked instruction are due to the decreased cognitive load that learners experience during instruction. Therefore, learners perceive that this format is more effective for learning simply because it is less taxing on their working memories. Blocked instruction may help learners process similarities within categories whereas interleaving may facilitate processing between categories (Birnbaum et al., 2012; Bjork & Bjork, 2011). The ability to process between categories is a crucial skill in clinical practice because nurses and other health care professionals are expected to differentiate between physical assessment findings. The ability to understand the difference between, for example, a systolic and a diastolic murmur, rather than the ability to recognize that Patient A’s wheeze is different from Patient B’s wheeze, is important component to patient assessment and management. As Bjork (1994) stated, the “conditions that introduce difficulties for the learner – and appear to slow the rate of the learning – can enhance long-term retention and transfer … [and are considered] ‘desirable difficulties’”. Indeed, increasing germane load through mixed instruction does result in an increased working memory load. However, the auscultation test performance results overall and for near transfer support the results in the literature that this approach facilitates inductive learning, even if these beneficial effects might not
be perceived by the learners as such during instruction.

**Conclusion**

The three studies reported above demonstrate the use of cognitive load theory to guide and evaluate instructional design through manipulation of intrinsic, extraneous, and germane load variables in cardiac and respiratory auscultation skill development. While the study incorporating multiple representations of diagnoses suggested that this approach contributed to extraneous cognitive load for the novice learner, the findings of interleaving instruction support the benefits of incorporating a mixed instructional approach for auditory instruction to develop auscultation skill in health professional students.
References


Chapter 4

Summary and Conclusion

This thesis explored Cognitive Load Theory (CLT) as a framework instructional design for health professions education research. The clinical focus of these studies was on cardiac and respiratory auscultation and physical assessment skill development in senior-level undergraduate nursing students. The studies compared CLT and context-based learning environments and explored the effects of manipulating cognitive load variables on auscultation performance outcomes. Applying CLT in health professions education research has been advocated and the research questions and approaches taken in this thesis are consistent with these recommendations (Mayer, 2010; Van Merriënboer & Sweller, 2010).

Chapter 2 highlighted the need to consider extraneous cognitive load variables in instructional design and of aligning instruction objectives with appropriate tools for instruction. Teaching health professions students in authentic learning environments is important and the results from this study do not discourage the importance of, and need for, students to learn within future practice contexts. The ultimate goal of health professions education is to prepare learners for future clinical practice and this necessarily includes instruction in authentic, or context-based, environments. Indeed, non-clinician cognitive and education psychology researchers have stressed the importance of “external validity” in their studies such that explorations around CLT are situated within actual learning contexts and with practical learning applications (Bjork & Linn, 2006; de Jong, 2010;
Therefore, the application of CLT or the use of context-based learning environments do not fall on opposite ends of a instructional design spectrum, and the application of one model does not preclude consideration and integration of the other. Rather, the results of this study emphasize that CLT variables are important to instructional design and could be used to create effective learning environments, depending on the learning goals and the level of learner. The novice learners used in these studies most likely benefit from a very different approach to instruction than the more expert learner. These learner differences were not explored in this thesis.

Chapter 3 focused on manipulating cognitive load variables and evaluating its effects on auscultation test performance. The most interesting findings from this series of studies was the effects of interleaving for auscultation test performance, and these study findings support its continued use for future instruction with novice learners around auscultation skill development. While this chapter offered an initial exploration in health professions education with the manipulation of the cognitive load variables, there are many more paths to explore when using CLT for instructional design and research.

One issue is the lack of clear definitions of the cognitive load variables. As was found in the studies conducted in this chapter, an approach in the literature that has facilitated germane processing (multiple representations) was not found beneficial to learning auscultation sounds in this study. These findings may have been due to multiple reasons such as poor study design or the novice level of learner. Without the cognitive load variables being clearly defined from the outset,
The ability for hypothesis testing is limited and interpretation of whether a design feature contributes to germane load or extraneous load is necessarily post hoc.

A framework that has expanded on CLT and includes consideration of learner characteristics has been proposed (Schnotz & Kürschner, 2007). This framework involves identifying learner expertise and tailoring instruction to the optimum level of difficult for the learner, or the “zone of proximal development”. These approaches have been advocated, not only in the cognitive psychology literature, but in motor learning as well and has alternatively been termed the “challenge point framework” (Guadagnoli & Lee, 2004; Guadagnoli, Morin, & Dubrowski, 2012). Such frameworks are similar in that they require consideration of learner characteristics in addition to determining the difficulty of the learning goal and the features of instruction that inhibit or facilitate learning of that goal. Continuing to conduct studies that explore instructional design using these frameworks for health professions education research, and particularly in nursing education research, allow for research questions to be situated at the nexus between the science of learning and the science of instruction, where a research theory is being tested within an authentic learning context (Mayer, 2008). The questions in this thesis were guided by these principles of learning and instruction and offered an initial exploration of cognitive load theory in the instructional design for cardiac and respiratory auscultation skill development for novice undergraduate nursing students.
References


Appendix

Development of a Clinical Performance Assessment Scale: Evaluation of Reliability and Validity

Background

The health professions education literature has explored the development and use of measurement scales to assess health professional student clinical performance (Martin et al., 1997; Vassiliou et al., 2005; Wass, Van der Vleuten, Shatzer, & Jones, 2001). Health professions educators seek to measure learning outcomes in students and to evaluate the effectiveness of education interventions. Measurement scales must provide sufficient reliability and validity when assessing these intended learning outcomes and the scales are constructed so that performance of the learners can be evaluated through objective criteria. These criteria have usually incorporated one or both common formats of measurement: the checklist and the global rating.

Checklists refer to an evaluator assessing a student’s performance by identifying individual actions or behaviors demonstrated as “done/not done” and students are given points for performing particular behaviors or actions and zero points if the behavior or action is not demonstrated. These points are then added together to produce an overall performance score. Global ratings, on the other hand, are based on evaluators providing an overall assessment of student performance on a particular domain based on an overall impression of student performance within that domain. Global ratings frequently use a Likert scale rating approach, usually involving either a 5-point or 7-point scale, and depend on the evaluator’s judgment; the evaluator is typically one who has expertise regarding the particular domain on which the student is being evaluated.
The literature has demonstrated that there is improved reliability and validity of global rating scales over checklists when assessing clinical skills performance (Regehr, MacRae, Reznick, & Szalay, 1998), meaning that while novice and expert practitioners could be differentiated with respect to their global rating scores, the checklists were not as effective in differentiating between the performance of the novice versus the expert. Therefore, while the novice performer might be able to demonstrate the individual steps of a particular skill or practice domain, the expert practitioner may miss particular steps on a checklist-based assessment yet still demonstrate superior overall clinical proficiency or diagnostic accuracy.

While there is a larger body of literature in medical education and in clinical fields such as surgery and anesthesia around the development and use of measurement scales to evaluate clinical performance, less literature is available regarding the use of measurement scales for evaluating nursing students. In particular, there is no known measurement scale to evaluate students’ pediatric cardiac and respiratory physical assessment skills.

**Study Objectives**

The purpose of this study was to develop and appraise the psychometric properties of a measurement scale evaluating pediatric cardiac and respiratory physical assessment skills of nursing students and non-specialist practicing nurses. Additionally, to evaluate construct validity of the scale, the hypothesis was that practicing nurses would achieve higher scores on the global rating scores than the student nurses.

**Methods**
Participants. Third level undergraduate nursing students (n = 12) in the undergraduate BScN program at McMaster University and pediatric nurses (n = 6) at McMaster Children’s Hospital were recruited to participate in this study. Participants were recruited between January – April 2008. Neither group of participants had previous practice or exposure to the patient simulators or to the simulation scenarios.

Scale development. A Likert scale assessing Global performance and Checklist items with dichotomous response variables (specific task completed/not completed) were used. Both the Checklist and the Global rating scales were used to ascertain whether findings would be consistent with what has been reported in the literature.

This measurement scale focused on demonstration of appropriate cardiac and respiratory physical assessment skills by a non-specialist nurse. The exam evaluated only the visual and auditory components of the physical exam and did not assess palpation (e.g. increased precordial activity) or percussion (e.g. consolidation in lung fields) skills. See Figure 1 for the criteria used for the Checklist and Global ratings.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Check if completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrates appropriate landmarks on patient</td>
<td></td>
</tr>
<tr>
<td>Demonstrates consistent comparison of left and right sided breath sounds</td>
<td></td>
</tr>
<tr>
<td>Listens for a full patient breath at each location</td>
<td></td>
</tr>
<tr>
<td>Counts heart rate</td>
<td></td>
</tr>
<tr>
<td>Counts respiratory rate</td>
<td></td>
</tr>
<tr>
<td>Demonstrates systematic approach to physical exam</td>
<td></td>
</tr>
<tr>
<td>Verbalizes observations of patient’s overall clinical presentation</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1. Physical Assessment Measurement Scale**
The Checklist for the physical assessment domain contained seven dichotomous response variables that represented a set of clinically important nursing assessment behaviors and actions. The Global rating had one 7-point Likert scale used for evaluation of participant performance.

**Simulated clinical scenarios.** Two simulated scenarios were used for this study, using human patient simulators and clinical simulation rooms. In Scenario 1, an 8-month old infant was admitted for bronchiolitis. In Scenario 2, a child was admitted with a right pneumothorax. SimBaby (Laerdal) and PediaSim (METI) were used for the infant and child scenarios respectively and the clinical simulation rooms were set up to replicate actual pediatric patients’ rooms. Every participant performed a 5-minute assessment on each simulated patient scenario. The participants were instructed to verbalize their thoughts and clinical assessments as they performed the physical exams.

All physical examinations, for both students and nurses, were video recorded. Each participant’s physical assessment recording was shuffled and compiled into a DVD for two raters to evaluate at a later date. Two expert clinicians who were unfamiliar with the participants involved in this study independently reviewed the participant scenarios and completed the Checklists and Global rating scales.

Each participant completed two scenarios and received independent ratings by two outside raters/observers. Checklist scores for each scenario were calculated for each participant by summing all checklist items; Global rating scores were assigned by each outside rater. Therefore, each participant had a total of 8 scores: Infant Checklist (Rater 1, Rater 2); Child Checklist (Rater 1, Rater 2); Infant Global (Rater 1, Rater 2); Child
Global (Rater 1, Rater 2). These scores were used to calculate the inter-scenario
generalizability and the inter-rater generalizability coefficients.

**Research ethics.** The joint Research Ethics Board of McMaster University
Faculty of Health Sciences, Hamilton Health Sciences, and St. Joseph’s Healthcare
Hamilton approved this study. All participants provided written and informed consent to
participate in this study.

**Statistical analyses.** SPSS version 15.0 was used for data analysis. Scores for
each participant, averaged across observers, were calculated for the mean (SD) Total
Checklist and Global Rating Scores for the Infant and Child Scenarios. To calculate the
inter-observer and the inter-scenario generalizability coefficients, Variance Components
Analysis were used which partitioned the variances attributed to the participants, the
scenarios, the observers, and to the interactions between these components. One-way
ANOVA was used to compare Clinical Findings scores between students and nurses.
Alpha level for all tests (two-tailed) was set at 0.05.

**Results**

Undergraduate nursing students had limited pediatric clinical experience as they
were all third year students with no previous degree; practicing nurses had an average
clinical experience of 12 years (range = 6 to 18). Table 1 shows the mean scores (SD) of
the student and nurse ratings averaged across the two observers. Overall, pediatric nurses
received higher scores than students on both checklist scores and global ratings. This was
consistent across both the infant and child scenarios.
Table 1. Descriptive data from scale, averaged across observers

<table>
<thead>
<tr>
<th></th>
<th>Student score Mean (SD)</th>
<th>Mean score (SD), Nurse</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infant Scenario</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Checklist</td>
<td>3.77 (1.88)</td>
<td>6.17 (1.19)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Global rating</td>
<td>3.23 (1.88)</td>
<td>5.71 (1.03)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><strong>Child Scenario</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Checklist</td>
<td>3.75 (1.65)</td>
<td>4.75 (0.89)</td>
<td>0.024</td>
</tr>
<tr>
<td>Global rating</td>
<td>3.60 (1.74)</td>
<td>5.29 (1.12)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Generalizability coefficients.** Overall reliability was 0.52 for the Checklist and 0.40 for the Global ratings scale. Inter-rater reliability was 0.38 for the Checklist and 0.30 for the Global ratings. Reliability of the measurement scale across scenarios (Inter-scenario generalizability) was 0.47 for the Checklist and 0.74 for the Global ratings.

**Construct validity.** Practicing nurses consistently achieved higher scores than nursing students on the Checklist and Global rating scales for both Infant and Child scenarios. For the comparison between Student and Expert Nurse performance, mean of student = 3.59, and for expert = 5.48. F(1, 34) = 16.11, p < 0.001. The Global ratings between expert and student nurse for the Infant Scenario [F(1,17)= 4.25; p < 0.001] and Child Scenario [F(1,17)=3.04; p = 0.005] and the Total Checklist scores for the Infant Scenario [F(1,17) = 4.02; p < 0.001] and Child Scenario [F(1,17) = 2.36; p = 0.024] were all statistically significant.

**Discussion**

This measurement scale designed to evaluate pediatric physical examination skills demonstrated fair reliability and construct validity. The reliability coefficients generated from the study data were not uniformly consistent with the literature that indicates improved reliability and validity of global rating scales over checklists when assessing clinical skills performance (Regehr et al., 1998). The data from this study did not confirm
our hypothesis that checklist scores consistently would not discriminate between the student and nurse groups as has been found elsewhere in the literature (Martin et al., 1997); nurses performed significantly better on both the infant and child scenarios. By having raters complete the checklist and the global rating at the same time, there may have been a halo effect between the checklist and global ratings such that this difference between checklists and global ratings could not be replicated.

Limitations

One limitation of this study was that the raters were not sufficiently blinded to participant physical characteristics during their assessments of participant performance. Therefore, although both raters did not know any of the participants, and even though all participants were asked to dress in hospital uniforms for the study, several of the pediatric nurses looked older than many of the students. Practicing nurses may therefore have received higher scores on the Checklist and Global rating scales due to being perceived as older and more experienced.

The second limitation was that the human patient simulators used for the clinical scenarios did not have the capability to produce all the physical findings of an actual pediatric patient. Therefore the reliability and validity of this measurement scale could not necessarily be generalized to evaluation of students in other contexts or in evaluation of assessment skill on actual pediatric patients. The limitations of the simulators preclude broader generalizability of this measurement scale. For example, when listening to heart sounds, the participants were only able to hear sounds through the speaker located at the apex of the heart. Therefore, even if the participant demonstrated appropriate landmarks for cardiac auscultation, sounds could not be heard elsewhere in the precordium. The
same was true for the respiratory sounds as lung sounds could be auscultated on the right
and left sides generally, but not specifically within each of the separate lobes of the lungs.

The third limitation in the construction of the measurement scale was that there
was only one global rating scale for assessment performance. This could explain why the
generalizability coefficients were low; increasing the number of items on which the raters
could provide global ratings could improve the reliability of the measurement scale
(Streiner & Norman, 2008).

Conclusion

The purpose of this study was to determine the reliability and validity of a
measurement scale for evaluating pediatric cardiac and respiratory physical assessment
skills, particularly within a simulated pediatric clinical context. Findings suggest that this
measurement scale had fair reliability and validity and can be used to assess nursing
student performance on pediatric physical assessment within clinical simulation contexts.
References


