Manipulation of Cognitive Load in Simulation-Based Medical Education

By

Eric A. Bruder

Queen's University

A thesis submitted to the Graduate Program in Education in conformity with the requirements for the Degree of Master of Education

Queen's University

Kingston, Ontario, Canada

Final (QSpace) submission March, 2018

Copyright © Eric A. Bruder, 2018

Abstract

In order to maximize the educational benefits of simulation based medical education (SBME) on an individual level, medical educators can manipulate simulation scenarios in real time, using the fundamentals of cognitive load theory. This approach requires accurate measurement of cognitive load. In this study, we investigate whether self-reported assessment of cognitive load using a visual analogue scale (VAS), and measurement of the galvanic skin response (GSR) are able to reliably and accurately measure cognitive load during SBME. Five separate experimental conditions are reported, and self-reported VAS and GSR measurements of cognitive load are compared before and after interventions designed to either increase or decrease cognitive load. Twenty-one junior postgraduate medical trainees participated in this study. Data were analyzed using repeated measures ANCOVA. VAS and GSR were highly correlated with each other, and with increasing numbers of recall items in a standard visual spatial memory task. While VAS was more sensitive for discriminating between varying levels of cognitive load during the visual spatial memory task, both VAS and GSR were able to do so. In two simulation scenarios designed to increase cognitive load, increases in VAS measures of cognitive load were statistically significant, while GSR measures were not. GSR decreased significantly in an exercise designed to decrease cognitive load, while VAS did not. Introduction of the worked examples effect in final simulation exercise resulted in statistically significant reductions in both VAS and GSR measures of cognitive load. Limitations and applications of the results are discussed.

Keywords: Cognitive load, simulation based medical education, galvanic skin response, worked examples effect.

Co-Authorship

None

Acknowledgements

The collection of data would not have been possible without the generous donation of personal time, simulation laboratory time and dedication to this project, by Kim Garrison (Operations Manager, Queen's University Clinical Simulation Centre), Jeremy Babcock (Lead Technical Lab Assistant, Queen's University Clinical Simulation Centre), Loren Fleming (Technical Lab Assistant, Queen's University Clinical Simulation Centre). I could not have collected the data for this project without the help of my colleagues, who volunteered their time to play the roles of confederates during the simulation exercises. Thank-you Jane Lewis, RN, Logan Burns, RN, Hailey Williams, RN, Christine Figeys, and Dr's Heather White, Emily House, Mackenzie Howatt, Kristen Weersink, Sharleen Hoffe, and Matt White.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER ONE: Introduction

Simulation based medical education (SBME) is a modern answer to an ancient question: how to balance patient safety with the educational needs of medical trainees. Without clinical exposure to real world patients, medical learners are unable to develop the skills necessary to provide care to the sick or dying. Until recently, the answer was to allow inexperienced learners to practice with supervision (of varying intensity) on real patients. Expertise was difficult to attain in this model, as the relative frequency of these experiences was low. As a result, sick and dying patients, in their time of greatest need, would be exposed to care providers with little experience, who lacked the opportunity to achieve mastery. It is no longer acceptable for inexperienced practitioners, such as medical students or resident physicians, to "practice" skills on patients without first demonstrating competence (Kohn, Corrigan, & Donaldson, 2000).

In response to this imperative, SBME was developed, and allows medical educators to expose learners to critically ill, simulated, patients in order to gain mastery before unleashing them onto the wards. Simulated clinical scenarios enable learners to explore their practice patterns in a safe environment, where mistakes can be made without harming patients or their own sense of self efficacy. It also allows for repetitive exposure to rare clinical entities, which traditional practice does not. Simulation can be used to teach psychomotor skills, as well as complex decision-making, leadership, and crisis resource management.

The use of SBME is resource intensive. In order to provide a small group of learners (typically 5) with a simulated clinical experience, a faculty expert debriefer, a technician, and one or two paid actors, are required. Frequently a second faculty member is also involved. In the face of the costs involved in the delivery of an SBME curriculum, we as medical educators are tasked with minimizing the cost to benefit ratio of this educational intervention, which includes the identification of the components of SBME that lead to enhanced learning and clinical performance.

One approach is the application of cognitive load theory in the development and execution of simulated clinical experiences (Fraser, Ayres, & Sweller, 2015). Cognitive load theory describes the mental architecture and processes used to affect learning, and educational design. This is particularly useful in the simulation of complex medical resuscitation exercises and crisis resource management. Currently, simulation educators design scenarios aimed at the average participant. This approach of designing simulation experiences to the group average provides a disservice to those learners who stray from the mean. In any given group, some learners will be advanced, and if the simulation exercise is not sufficiently challenging, these learners will gain little from the experience. Likewise, weaker learners may become overwhelmed by the complexity of the task, disengage, and gain very little from their experiences.

Ultimately, if we endeavor to customize the simulated experience to individual learners, effectively ensuring that all individuals are operating within their zone of proximal development, then we shall maximize the educational efficiency of SBME. The challenge becomes one of measuring and interpreting their cognitive load to ensure they are employing sufficient cognitive resources to maximize learning, without approaching cognitive overload and disengagement. If we are able to measure the learner's cognitive load, then interventions based on cognitive load theory should allow us to manipulate the simulation scenario in order to maximize the educational experience in real time. The aim of this study is to investigate whether we can accurately and reliably measure cognitive load in a complex simulation laboratory environment.

Chapter 2 reviews the basis of SBME, cognitive load theory (CLT), and explores the measurement of cognitive load through various modalities in attempts to alter cognitive load in a dynamic environment during simulated medical experiences. Chapter 3 describes the methods, with which we will both alter, and measure, cognitive load during simulated clinical exercises using high fidelity simulation. In chapter 4, we will compare two measures of cognitive load (self-reported scales, and the galvanic skin response) during a controlled visual spatial memory task, and four simulated resuscitation exercises. We will analyze our data using repeated measures ANCOVA. In our final chapter, we will discuss the relevance of our findings, some limitations of our work, and suggest future directions for measuring cognitive load in active and dynamic environments.

CHAPTER TWO: Literature Review

In this chapter, we explore the rationale for simulation-based medical education plays in the training of physicians. The environment of simulation based medical education is active and dynamic, and in order to achieve the best educational value from the approach, many medical educators employ educational techniques derived from cognitive load theory. After reviewing cognitive load theory, we will review the physiological basis and techniques of measuring cognitive load so that we may understand how changes in cognitive load are quantified in this study.

Simulation Based Medical Education (SBME)

Despite an explosion in popularity and frequency of use, little is known about whether SBME leads to better learning or bedside patient care. While individual studies of SBME have demonstrated superior simulation based assessment outcomes as well as significant improvements to patient care (W. C. McGaghie, Issenberg, Barsuk, & Wayne, 2014), when meta-analyzed, no significant benefits to patient oriented outcomes are demonstrated (Brydges, Hatala, Zendejas, Erwin, & Cook, 2015; Zendejas, Brydges, Wang, & Cook, 2013). In their meta-analysis of 33 studies (Brydges, et al., 2015), including a total of 920 medical trainees, only 5 studies reported patient-oriented outcomes. Only one of these studies demonstrated a significant improvement in patient-oriented outcomes, and their overall effect size was small and their confidence intervals crossed the line of no effect. These studies also demonstrated a large degree of heterogeneity. Similar results were seen in a meta-analysis comparing SBME to traditional medical education (Zendejas et al., 2013). In this meta-analysis of 34 studies (1694 participants), the pooled effect size comparing SBME to standard medical training was small, the confidence intervals crossed the line of no effect, and the heterogeneity was large. When SBME was compared to no training (i.e. no comparison intervention, or SBME as the sole educational intervention), a small to moderate effect size in favour of SBME was observed, although the heterogeneity of the studies remained large, limiting the validity of their results (Zendejas et al., 2013).

In a systematic review of simulation research methodology and quality, Schaefer et al. (2011) identified significant methodological challenges in the SBME literature making it impossible to draw conclusions about its efficacy. One single reviewer screened over 4000 articles, before distributing 221 articles between 7 reviewers for assessment of inclusion eligibility. Their search results were not reported according to the PRISMA methodology (Moher, Liberati, Tetzlaff, & Altman, 2009), and they did not perform an analysis of inter-rater reliability among their 7 reviewers. They identified 221 articles in their search. Seventy percent of the articles did not describe the validity of their simulated assessment tool, and 90% failed to report reliability. Only 112 reported outcomes in favour of SBME (Schaefer et al., 2011).

Another systematic review compared high fidelity simulation to other methods of nursing education (Cant & Cooper, 2010). Twelve studies were included out of an initial 2019 studies identified in their electronic search. Four of 9 studies designed to demonstrate knowledge improvement showed statistically significant increases in knowledge in the simulation groups compared to controls. Five of 11 studies of critical thinking reported statistically significant increases in self-reported confidence in the ability to make clinical decisions compared to controls, however when validated scales of critical thinking (e.g. the California Critical Thinking Disposition Inventory) were reported, there was no observed difference in critical thinking. Only one randomized controlled trial showed increased critical thinking using a standardized assessment (Cant & Cooper, 2010).

In a systematic review of SBME, feedback, repetitive practice, appropriate difficulty level, and individualized learning were identified to be among the most important factors in SBME. Other factors include integration of simulation into the medical curriculum, the use of multiple learning strategies, clinical variation, environmental control, clear objective benchmarks, and simulator validity (Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005).

A meta-analysis involving 633 learners from internal medicine, surgery, and emergency medicine training programs demonstrated that simulation based medical education, with deliberate practice, improved advanced cardiac life support (ACLS) skills, laparoscopic surgical skills, central venous catheter insertion, cardiac auscultation (listening to heart sounds on physical examination), and thoracentesis (a procedure in which a needle is inserted into the thorax to remove fluid from around the lung) (McGaghie, Issenberg, Cohen, Barsuk, & Wayne, 2011). Overall the effect size of the interventions was classified as large (random effects meta-analysis, overall effect size 0.71, 95%CI 0.65-0.76), leading the authors to conclude that “there is no doubt that SBME is superior to traditional clinical education for acquisition of a wide range of medical skills” (McGaghie, et al., 2011, p. 708).

In another study, just 4 hours of simulation training with deliberate practice and feedback, on central venous access, emergency medicine and internal medicine residents were able to achieve statistically significant reductions in attempts at central line insertion, catheter adjustments, and arterial puncture (a complication of the procedure) compared to peers trained with traditional methods (Barsuk, McGaghie, Cohen, O’Leary, & Wayne, 2009). These SBME trained residents also demonstrated statistically significant improvements in success rates on real ICU patients. While this study was an observational cohort study before and after a change in educational technique, and not an experimental design, the strengths of their findings lend significant support to SBME’s efficacy at teaching psychomotor skills.

Very little SBME training is required to show improvements in some cases. In a study of 21 emergency medicine residents, only 20 minutes of a video tutorial followed by 10 minutes of self-guided practice improved procedure times and efficiency during simulated fiber-optic intubation (Binstadt, Donner, Nelson, Flottemesch, & Hegarty, 2008). These improvements were only observed in 2 out of 3 of their outcome assessment scenarios, and the lack of a control group, small sample size and limited educational intervention challenge the validity of their results. Issenberg et al. (2002) found that after only 10 hours (5, two hour sessions), medical students and residents were able to improve their bedside physical exam skills 4.3 to 6.4 fold above their baseline abilities in a pre-test, post-test designed study. No such difference was observed in a control group (Issenberg et al., 2002).

Advanced cardiac life support (ACLS) is a standardized training program that teaches physicians how to treat life threatening cardiac dysrhythmias. It is the standard resuscitation training obtained by the majority of medical trainees, and forms the basis of the approach to life-threatening conditions for most physicians. ACLS training has also adopted SBME in its courses. When a cohort of ACLS learners participating in a course using high fidelity SBME was compared to a cohort using low fidelity simulation (SBME vs standard training), statistically significant differences in confidence, knowledge, and appropriate treatment decisions were observed using expert rater video review outcomes (Rodgers, Securro, & Pauley, 2009). In a randomized control trial of 38 internal medicine residents undergoing ACLS training in a cross-over design, SBME led to a 38% difference in outcome scores compared to the control group. The control group demonstrated similar improvements in performance following cross-over (Wayne et al., 2005). Another pre-test, post-test designed study, without a control group, involving 41 internal medicine residents with little to no ACLS experience, showed that 4, 2-hour ACLS education sessions utilizing SBME led to an absolute increase in their post-test scores of 24% (Wayne et al., 2006).

Resuscitation skills are frequently taught using high fidelity simulation. The foundations of cognitive load theory provide a sound rationale for the design and implementation of SBME (Fraser et al., 2015). Further understanding of cognitive load theory will allow SBME educators to tailor the simulation exercises to the capacity of the individual learner, so that each learner will reap the maximal benefit from each training session.

Cognitive Load Theory

A common model of our cognitive processing architecture is cognitive load theory (Sweller, Ayres, & Kalyuga, 2011). Cognitive load theory outlines the framework within which our working memory handles and manipulates information ultimately affecting change in our long-term memory. Educators use cognitive load theory to optimize the learning environment and material (Chandler & Sweller, 1991; Kalyuga, 1999; Mayer & Moreno, 2003; Paas, Renkl, & Sweller, 2003). This is of particular importance in learning environments where learners are presented with information in a variety of media, such as computer training, multimedia presentations, or simulated medical training. Techniques to decrease excessive cognitive load, which may overwhelm learners, are derived from this theory as well (Mayer & Moreno, 2003; Paas, Renkl, et al., 2003; Paas & Van Merrienboer, 1994a; van Merrienboer, Kirschner, & Kester, 2003). Cognitive load theory describes our cognitive architecture in terms of discrete mental constructs including long-term memory, the sensory register, working memory, and the over-riding central executive.

Long-Term Memory and Learning. The goal of the teacher and learner is to effect change in the learner's long-term memory. This change is considered learning. Long-term memory is thought to be limitless, and units of knowledge are stored in long-term memory as schemas. Data are manipulated in working memory before being stored in long-term memory as schemas. Similarly working memory can recall schemas, as single units or concepts, from long-term memory in order to minimize the cognitive load placed on working memory (Ericsson & Kintsch, 1995; Sweller et al., 2011). As expertise increases, so does the volume and complexity of the individuals' schemas, allowing experts to handle vast amounts of complex information as a single unit.

Sensory Register, Channels, and Attention. The sensory register is the cognitive entity that must handle the various stimuli presented to the learner. The sensory register filters the stimuli from our surroundings and determines what information is transferred from our environment into working memory. It has been proposed that there are two "channels" within the sensory register, auditory and visual, and that each channel can only register one stimulus at a time, while other stimuli must be ignored (Mayer & Moreno, 2003). In contrast to this description, consider that we have five senses. Although not specifically targeted by an instructional task, all of our senses must be attended to or ignored, and any sensory input that is not germane to our learning tasks may detract from it. When more than one stimulus is encountered per channel, or sense, attention must intermittently shift to and from each stimulus. As an example, it is not possible to read two texts at the same time, because you cannot see two texts at once. Consider the learning environment where extremes of temperature, background noise, or offensive odors exist (e.g. clinical medicine). In order to concentrate on the task at hand, learners must exclude these stimuli from their working memory in order to succeed. Many times in medicine, we use our sense of touch, smell, and (historically) taste to gather clinical information. These senses must be acknowledged in the medical learning environment and are an important data set to the clinician.

Working Memory. Working memory is the realm of our cognitive architecture that manipulates data. Working memory receives information from the sensory register, pulls schemas from long-term memory that are relevant to the sensory cues, digests and refines them, and then stores new or modified schemas back into long-term memory (Sweller et al., 2011).

***Fixed and Finite*.** Traditionally, working memory is thought to have a fixed and finite limit of seven (plus or minus two) “chunks” of information. The exact capacity of working memory remains a mystery and may be as few as three to four pieces of information. Working memory is also thought to have a temporal limit of only 20 seconds for novel information (Sweller et al., 2011). Situations that increase cognitive load decrease the available capacity of working memory. Working memory holds three types of cognitive load: germane, intrinsic, and extraneous. While the overall capacity of working memory is fixed, the proportions of germane, intrinsic, or extraneous load vary. In order to increase one proportion of cognitive load, another must decrease, given the finite capacity of working memory. That to say, in order to increase germane cognitive load, intrinsic cognitive load, or extraneous cognitive load (or both) must decrease (Paas, Tuovinen, Tabbers, & VanGerven, 2003; Sweller et al., 2011).

Germane Cognitive Load. Germane cognitive load refers to all the mental processes that are central to learning (Sweller et al., 2011). These processes may include assimilation, accommodation, representation holding, repetition, or memory processes. As educators, we must strive to maximize the germane cognitive load in our learners, as this will optimize learning. We can coach germane cognitive load by giving our students metacognitive strategies for learning.

Intrinsic Cognitive Load. Intrinsic cognitive load is a function of the difficulty of the content of the learning task for individual learners and their level of expertise (Sweller et al., 2011). This is specific to each particular learning task, and cannot be manipulated by the instructor, as it is an integral characteristic of the task and learner. The more difficult the task, or the less skilled the learner, the higher the intrinsic cognitive load. As instructors, we can design our learning tasks to suit our learners such that the intrinsic cognitive load is appropriate for their level of experience and mastery, but once the task is set, the intrinsic cognitive load is fixed.

Extraneous Cognitive Load. Extraneous cognitive load represents all the cognitive processes that are undertaken that do not lead to learning or the successful completion of a task (Sweller et al., 2011). An extreme example would be number recall, or numeric representation holding, when the learning task at hand focuses on literature interpretation. A common example in medical simulation is having learners recall details about drug pharmacology and dosing, when the goal of the learning scenario is effective execution of CPR skills. Extraneous cognitive load can often be manipulated by the instructor, and should be minimized in order that the proportion of working memory dedicated to germane and intrinsic cognitive load is maximized (Chandler & Sweller, 1991; Fraser et al., 2015; Sweller et al., 2011). Depending on the objectives of the learning task at hand, some processes may impose extraneous cognitive load in one circumstance, while imposing germane cognitive load in others. In the example above, recalling drug pharmacology may be germane, if the outcome of interest is drug selection and dosing, and not the mechanical skill of CPR.

Element Interactivity. Element interactivity is an estimate of task difficulty or complexity, and its overall magnitude can be difficult to estimate, as this estimate depends on the novelty of the information as well as the learner's previous experience and knowledge. An "element" is any bit of information manipulated within working memory. A schema recalled from long term memory can be utilized as a single element, and therefore, experts can manipulate vast volumes of information in their working memory, for concepts with which they are well acquainted (Sweller et al., 2011). Novel information, however, must be handled as discrete elements since no schema pre-exists to consolidate data into a single unit (O. Chen, Kalyuga, & Sweller, 2015). The addition of elements manipulated by working memory increases the cognitive load of a learner. These elements, depending on their nature, may impart intrinsic, germane, or extraneous cognitive load on the learner. Tasks with excessively high levels of intrinsic cognitive load adversely affect the execution of germane cognitive load processes. Learners can manipulate their intrinsic cognitive load by recalling schemas from their long term memory. In effect, this allows them to manipulate a schema as a single unit of information, instead of using all the sub-components individually. This is one of the metacognitive skills we can coach our learners to use. As an example, medical trainees are often overwhelmed by the complexity of simulated resuscitation exercises, but have very clear schematic or algorithmic representations of the four main causes of cardiac arrest. When this cognitive overload is identified, instructors can have the learner refocus their efforts within one of the four algorithmic schemas, thereby decreasing the intrinsic cognitive load.

Emotion. Learners enter any environment in some form of emotional state. These emotions may change the learner's experience of the learning tasks, and the learning tasks may affect this relatively neglected entity. Changes in emotional state can have both positive and negative effects on learning, depending on the learner's response to these emotions. Emotions can be viewed as additional elements that add to the total cognitive load of a learner (Fraser et al., 2015). Both positive and negative emotions impart a cognitive load on a learner. In a study of 1st year medical students, emotions were shown to impart an extraneous cognitive load that decreased the students' ability to identify abnormal heart sounds (Fraser et al., 2012). We, as educators must be aware of the effect that these have on our learners. In high fidelity simulation, we seek to engage learners' emotions, as these are of critical importance when dealing with real patients.

Sympathetic Response to Stress in Simulation-Based Medical Education

When faced with stress, our body produces a "fight or flight", or adrenergic, response through activation of the sympathetic nervous system (Kreibig, 2010; Landsberg & Young, 2001). This involuntary response is activated by stress that is either real or imagined. Cognitive effort is a form of stress perceived by our body that results in the activation of the sympathetic response and the release of cortisol (Bohnen, Houx, Nicolson, & Jolles, 1990; Kirschbaum & Hellhammer, 1994; Kreibig, 2010; Schreinicke, Hinz, Kratzsch, Hüber, & Voigt, 1990). Previous SBME evaluation of the sympathetic response shows that more complex and stressful simulation exercises result in significantly higher salivary cortisol levels (Leblanc et al., 2012). The total cognitive load of a task is assumed to be proportional to the perceived emotional stress, and resulting adrenergic response (Jezova, Makatsori, Duncko, Moncek, & Jakubek, 2004; Kreibig, 2010). We will discuss a brief synopsis of the adrenergic response, in order to frame our impending discussion of its measurement.

Physiological Response to Stress

In response to stress, excitatory neurotransmitters are released from synapses, typically in the form of nor-epinephrine (NE), which is also known as nor-adrenalin. This release of NE has varying effects on different organ systems, which ultimately produce the adrenergic response (Landsberg & Young, 2001). Norepinephrine and epinephrine are together known as catecholamines.

Adrenal Glands. The adrenal glands, situated on top of each kidney, produce hormones that act to augment the adrenergic response. The adrenal glands produce adrenalin (epinephrine), cortisol, and other hormones which all serve to promote physiologic arousal, vascular tone, fluid retention, and the body's metabolic ability to respond to the adrenergic response (Landsberg & Young, 2001).

Heart and Blood Vessels. Beta-type adrenergic receptors in the heart respond to catecholamines by an increase in both the rate at which the heart beats, and the force with which it contracts (Landsberg & Young, 2001). When combined with the vasoconstriction changes in the skin described below, this results in an increase in blood pressure.

Skin. Alpha-type adrenergic receptors in the blood vessels of the skin result in a decrease in blood flow to the skin through a process called vasoconstriction. This results in greater resistance to blood flow, thereby redirecting blood to areas of the body where it is needed more (Landsberg & Young, 2001). The skin also produces more sweat from the sweat glands, in an effort to maximize evaporative cooling in anticipation of physical activity that leads to hyperthermia.

Pupils. Stimulation of the ciliary muscles in the iris of the eye, via the oculomotor nerve, leads to their contraction. This is also mediated by alpha-type adrenergic receptors. This leads to dilation of the pupils (Landsberg & Young, 2001).

Other various metabolic responses to stress such as muscle tension, enhanced cellular metabolism, and gastrointestinal motility are not directly measurable in the context of cognitive load measurement, and will not be discussed further.

Measuring Cognitive Load

It stands to reason, that if cognitive load is so entwined in our learning behaviour, so critical to the design of our learning environment, and so important to the implementation of our learning tasks, then measuring cognitive load must be a valuable tool to manipulate our teaching and learning. To date, we have not been able to differentiate the various components of cognitive load when recording surrogate markers of cognitive load. Instead different metrics are assumed to correlate to the total cognitive load experienced by the learner. When we measure surrogates of cognitive load, we are actually measuring the body's adrenergic response to the emotional stress imposed by the learning task. Various methods have been devised to measure cognitive load. These methods can be either biometric (measurement of the body's physiologic responses to the stimulus), task performance based (where the performance of concurrent tasks assumes the ability to differentiate the cognitive load imposed by each task), or by introspection.

Self-Report Scales. Self-report scales of cognitive effort have been shown to be non-invasive, sensitive enough to differentiate cognitive load levels within a task (Paas, Van Merrienboer, & Adam, 1994) and to correlate well with task difficulty (Gopher & Braune, 1984). A common tool uses a 9-point Likert scale and requires subjects to reflect on their own cognitive effort applied to a task. When studied by Pass, et al. (1994), retrospective self-evaluation was an effective tool for assessing cognitive effort of subjects. In this report, the authors summarize the reliability and sensitivity of this self-reported measure of cognitive load, involving 42 participants tasked with completing 12 computer based statistical analyses (Paas, 1992), and a second study 60 participants solving geometric problems (Paas & Van Merrienboer, 1994b). This scale reports the cognitive effort put forward by subjects, and this is assumed to be a valid surrogate measure of cognitive load. In the first study Cronbach's alpha for the self-reported scale was 0.90, and this self-reflection was sensitive enough to differentiate between each task with varying degrees of difficulty using a one-way ANOVA (Paas, 1992). In the second study, Cronbach's alpha was 0.82, and the scale was again able to differentiate between varying levels of cognitive load using a repeated measures ANOVA (Paas & Van Merrienboer, 1994b). Self-report rating scales have been shown to be superior to pupillary and eye blinking measurements, task completion time, and performance accuracy in the measurement of cognitive load. In this study, the *F*-statistic for self-reported mental effort were greater than for pupil diameter, blink rate, and task completion time using repeated measures ANOVA across 5 classes of progressively more difficult arithmetic problems, and self-reported cognitive load was more accurate at differentiating between tasks with varying difficulty (S. Chen, Epps, & Chen, 2011). The applicability of this self-report tool during a dynamic educational intervention remains to be proven. In the midst of a simulated clinical scenario, the ability of learners to introspect on their cognitive effort may be obscured by other factors such as ongoing cognitive efforts related to the simulated resuscitation. Further, the act of self-reflection may impart an extraneous cognitive load on the learner. If introspection is done after the event, success in the task may affect how difficult it is judged to have been, and limitations in the temporal capacity of memory may negatively impact the ability to recall information accurately. In light of these limitations of self-reporting cognitive load, we seek out other measures of cognitive load that are immune to these confounders. The biometric measurement of cognitive load can be applied to dynamic environments and measure cognitive load without interacting with the learner or the environment.

Heart Rate & Blood Pressure. Perhaps the most accessible biometric measures of sympathetic arousal, heart rate and blood pressure are insensitive modalities to measure cognitive load. Heart rate and blood pressure are affected by many environmental variables, and controlling for these variables in a learning situation is not feasible. Heart rate can be influenced, not only by emotional arousal, but also by temperature, fraction of inspired oxygen, physical fitness, altitude, and several medications. Although a seemingly intuitive approach to measuring physiological arousal in cognitive load testing, heart rate has been shown to be inferior to self-report scales, and while it can differentiate between periods of arousal and inactivity, it lacks the sensitivity to differentiate different levels of cognitive load within a task (Paas & Van Merrienboer, 1994b). In this study, 60 4th year technical school students had their cognitive load assessed by both the 9-point Likert scale of cognitive load, in addition to ECG analysis of heart rate variability while they completed various computer numerical control (CNC) and geometrical tasks. Analysis of heart rate variability using ANOVA and *t*-tests demonstrated that heart rate variability could differentiate between large differences in activity. This method could only differentiate between mental inactivity, and activity. This measure was not sensitive enough to differentiate between levels of cognitive load within a task. Self-reports of mental effort were able to differentiate between these within-task differences (Paas & Van Merrienboer, 1994b).

Pupilometry. Pupils respond to adrenergic stimulation by dilation. Pupillary dilation occurs quickly and reliably after changes in cognitive stress, and pupils return to normal shortly after the stimulus has subsided, making pupil measurement a promising metric for cognitive load. Pupilometry was first described in an experiment where subjects had their pupils photographed while performing mental arithmetic problems (Hess, 1964). In this experiment, it was shown that the pupil diameter increased as the multiplication problem was worked through, and returned to normal after the answer was given. Mean pupil diameter was larger for progressively more challenging tasks. The degree of pupillary dilation is proportional to the adrenergic response and the rate of dilation is proportional to the difficulty of the task at hand. This is true across a spectrum of cognitive tasks including memory recall, language interpretation, and complex problem solving tasks (Beatty, 1982). Pupillary measurements are able to accurately differentiate cognitive load within a complex task, and can differentiate the cognitive load experienced by novices and experts (Szulewski, Fernando, Baylis, & Howes, 2014; Szulewski, Roth, & Howes, 2015). Pupilometry correlates well with self-reported assessments of mental effort during medical test taking and mathematics problem solving (Szulewski, Gegenfurtner, Howes, Sivilotti, & van Merriënboer, 2016). Other eye-based measures of cognitive load currently being investigated include saccadic eye movement analysis, pupillary unrest, and gaze tracking (S. Chen & Epps, 2013; Marshall, 2007). Sophisticated pupilometry devices have been developed that allow subjects to interact with complex environments. These systems are non-invasive and provide excellent data sensitivity of task difficulty with accurate time data, but do not provide real time data required for instructional intervention in a simulated medical resuscitation. Pupillary size is also affected by ambient lighting. To overcome this limitation raw pupil diameter data can be manipulated to calculate an “index of cognitive activity” that is not affected by changes in ambient lighting (Marshall, 2007).

Evoked Potentials and Electroencephalogram. Electroencephalogram (EEG) data records surface scalp electrical activity originating in the brain. Several surface electrodes are used to identify which areas of the brain produce electrical discharges that occur when neurons are activated. In an experiment involving 5 young male volunteers, EEG was able to differentiate the cognitive load required for reading tasks of three levels of cognitive load (Zarjam, Epps, & Chen, 2011). EEG can provide real time data and is sensitive to subtle changes in cognitive load, when signal averaging techniques are used, and unique wave form analysis can distinguish between different cognitive processes (Antonenko, Paas, Grabner, & van Gog, 2010). Traditional EEG monitoring is cumbersome, and not practically applied to a physically active, or mobile subject. Electrical noise from skeletal muscle depolarization and the number of surface electrodes attached to the computer unit or analyzer make this an impractical choice for the measurement of real time cognitive load in medical simulation. Currently, industry has developed an EEG analyzer that fits onto the headband of eye tracking glasses, but this unit is currently cost prohibitive for use outside of supported research.

Functional Magnetic Resonance Imaging. Magnetic Resonance Imaging (MRI) creates computer-generated images based on energy released when magnetic fields are applied to a body tissue. Differences in water content in different tissues allow the computer to generate contrast between different tissues. When different areas of the brain are activated by cognitive, or other tasks, the water content changes, allowing MRI machines to map physiologically active areas of the brain. This produces an image of the “functioning” parts of the brain. While very detailed images and data can be created using this modality, it is not feasible to use this modality in real time cognitive load measurement in ambulatory subjects. Subjects who undergo fMRI study for cognitive function must do so while lying perfectly still for long periods of time. As such, we will not discuss this otherwise fascinating modality further.

Galvanic Skin Response. The electrical conductivity of the skin changes as the presence of sweat changes. As more sweat is produced, the resistance to electrical conduction decreases. This change in electrical conductivity is easily measured with minimally intrusive means, and can reliably differentiate between levels of cognitive effort within tasks including reading and mathematics tasks (Conway, Dick, Li, Wang, & Chen, 2013; Nourbakhsh, Wang, Chen, & Calvo, 2012), and complex computer based traffic control simulations (Shi, Ruiz, Taib, Choi, & Chen, 2007). The apparatus used to measure the galvanic skin response (GSR) consists of a pair of electrodes placed on the skin, and a recording unit that can be worn on a belt, or carried in a pocket. Data can be produced, transmitted, and interpreted in real time, allowing for educators or researchers to modify the learning task or environment in real time.

Given the portability, and minimally invasive nature, of GSR measurement, we will compare it to self-reported measures of cognitive load in a dynamic simulation-based medical education environment, using similar statistical methods as previous authors (S. Chen et al., 2011; Paas, 1992; Paas & Van Merrienboer, 1994b). We will utilize a standard visual spatial memory test to validate GSR and self-reported cognitive load measurement, before embarking on a series of simulated medical resuscitation exercises, where junior post-graduate medical trainees will employ their knowledge of ACLS. The details of which are described in the following chapter.

Research Question

In a population of junior postgraduate medical trainees, can changes in cognitive load experienced during simulation-based medical education be measured using the galvanic skin response and retrospective self-reports of cognitive load?

CHAPTER THREE: Method

Site of Study

This study was performed at the clinical simulation center at Queen's University. This facility is a modern high fidelity simulation center (http://meds.queensu.ca/education/simulation) utilized by emergency medicine and resuscitation faculty.

Ethics

The Queen's University health sciences research ethics board cleared this study (Appendix A).

Participants

A convenience sample of 21 participants was utilized. First and second year emergency medicine residents enrolled in the Royal College of Physicians of Canada (FRCPC) specialty training program at Queen's University were asked to participate in this study. A group of "off service" first and second year residents (i.e. not emergency medicine trainees) was also recruited during their emergency medicine rotation at Queen's. All junior residents were eligible for participation during their emergency medicine training rotation. Sixty-three residents were invited to participate in the study, and 21 did so. Participants were not selected based on training program, simulation experience, medical knowledge, gender, age, or ethnicity. None were excluded.

Data Analysis

Data were analyzed using SPSS (Version 24) and Microsoft Excel. Repeated measures ANCOVA were performed, with separate covariate analyses controlling for age, gender, trait anxiety, and baseline galvanic skin response.

Measures

Neulog GSR Logger (NUL-217) (https://neulog.com/gsr). The Neulog GSR (Galvanic Skin Response) unit measures the electrical conductivity of the skin between two points (measured in nano- or micro-Siemens). The recording unit was attached to the participants with two electrodes attached to two adjacent fingers on the non-dominant hand via Velcro straps. The sensor is pre-calibrated, and no further manipulation prior to data collection was required. The Neulog GSR unit is capable of transmitting data for experiments lasting from 1 second to 31 days, with a resolution of 10 nS, at a sampling rate of up to 100Hz. In our experiments, data were collected at a rate of one data point per second (1 Hz), for the duration of each task. Data from the GSR unit was transmitted by WiFi network to a laptop computer via the Neulog WIFI 201 module. Embedded software in the WiFi module allows the user to control GSR unit parameters, and exports data in tabular form to any spreadsheet application (i.e., Microsoft Excel).

Self-reported Scale of Cognitive Load. The 9 point Likert scale developed by Paas (Paas, & van Merriënboer, J. J. G., 1994b) was modified to a 100mm visual analogue scale (VAS) in order to produce continuous data. Participants received instruction on reporting mental effort and how to differentiate mental effort from task difficulty.

Tasks

After informed consent was obtained, participants completed a brief demographic survey and the Speilberger Trait Anxiety Inventory. The Speilberger Trait anxiety Inventory measures one’s baseline level of anxiety. The scale consists of 20 questions using a 4-point Likert scale. The lowest possible score is 20, and the highest possible score is 80. Trait anxiety increases proportional to the total score. (Barnes, Harp, & Jung, 2002). Participants were then administered a visual spatial working memory task and four simulation exercises over 1.5 hours.

Visual spatial working memory task. This task measured total cognitive load while participants solved an adapted visual spatial memory task (Swanson, 1995). This was intended to demonstrate that cognitive load measurements using GSR and self-reported scales are reproducible, compared to previously published work. Participants performed 11 items. In each item, a series of dots, positioned on a grid, was presented to the participants for 20 seconds on a projector screen. Participants were seated approximately 10 feet from an 8 foot x 5 foot screen. After 20 seconds, the visual stimulus was removed and the participants were given 20 seconds to recall and record the postilion of the dots on a blank grid of the same size. After recording their answers, the participants recorded their perceived mental effort, or total cognitive load, on a 100 mm visual analogue scale. This perceived mental effort captured the total cognitive load for the participants for each

40-second task. Each successive item became more difficult, ranging from 2 to 12 dots in each grid, and with the number of grid squares increasing from 4 to 45.

Simulation exercises. In each simulation exercise, the participant entered an empty simulation laboratory and sat quietly for two minutes (0 to 2.00) to acclimatize to the environment. A baseline GSR measurement was recorded from 1:45 to 2:00. At 2:00, the simulation exercises began, and the confederate actors entered the simulation laboratory. Two 15 second measurements (GSR and retroactive VAS) were taken at the beginning of each scenario, while the participant was interacting with the mannequin to obtain a clinical history and perform a physical examination. Microphones allowed the simulation technician to hear the questions posed by the participants, and speakers in the mannequin allowed the technician to provide a standardized scripted clinical history to the participants. These initial two measurements did not involve experimental interventions, and were separated by 15 seconds. Measurements of GSR and retroactive VAS (15 seconds) were taken before and after each experimental intervention described below. At the conclusion of each simulation exercise, the participants reviewed video of their performance during each of the 4 GSR measurements. Participants only reviewed the four time periods for which corresponding GSR data were collected. This retrospective review took place in a debriefing room that was separate from the simulation laboratory. The participants used the video to recall their performance and rated their total cognitive load on the 100mm VAS for each of the four 15 second portions of the exercise. All simulation exercises were pilot tested and refined prior to data collection. Each task had its own purpose, as described below:

Simulation Exercise 1: Increasing element interactivity. The clinical scenario simulated a young male with a massive pulmonary embolus (blood clot in the artery supplying the lung causing shock). There was very little ambiguity in the simulated clinical presentation. Participants obtained a clinical history and performed a physical examination of the mannequin. GSR and retrospective VAS measurements using video replay were obtained during this phase of the encounter. As the scenario progressed, the participants were provided with additional clinical information including an electrocardiogram (ECG), a chest X-ray, and changing vital signs with clinical deterioration. Each of these pieces of information added to the element interactivity experienced by the participants. At this stage, two more GSR and retrospective VAS measurements (3 and 4) of total cognitive load were recorded. The simulation exercise was terminated after this point. Figure 1 depicts anticipated cognitive load curve during the simulation exercise, and describes the participants' experiences and the time points where measurements of GSR and VAS were taken.

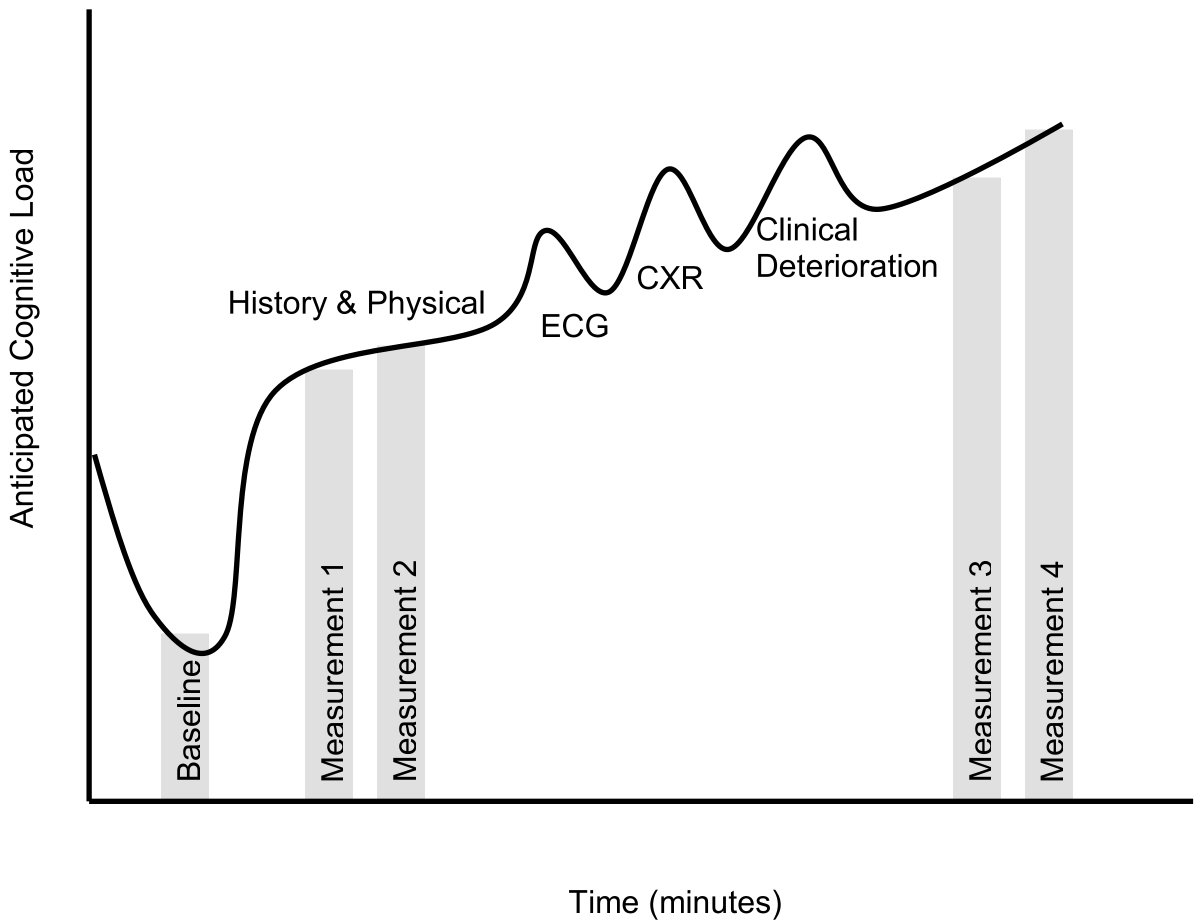

Figure 1. Schematic of anticipated cognitive load over time for simulation exercise 1. ECG=Electrocardiogram, CXR= Chest X-Ray.

Simulation Exercise 2: Unpack Schema: The simulated clinical scenario was one of pneumonia (bacterial infection of the lungs) leading to fulminant sepsis (overwhelming infection where the body is no longer able to control or regulate its response to the infection), and ultimately cardiac arrest in an elderly woman. There was very little ambiguity in the simulated clinical presentation. Participants obtained a clinical history and performed a physical examination of the mannequin. GSR and retrospective VAS measurements using video replay were obtained during this phase of the encounter. The simulated patient's condition deteriorated into cardiac arrest. After the participants began the resuscitation of the patient, a 15 second GSR and VAS measurement was taken before the experimental intervention. In this scenario, the element interactivity of a task was increased, by having the participants "unpack" previously stored schema regarding sepsis, thereby increasing the elements that must be handled in working memory to resuscitate the patient. A confederate medical student asked participants to describe the physiological mechanism of sepsis, a concept with which they are already familiar, and which was the focus of the resuscitation exercise. Post-intervention GSR and VAS measurements (15 seconds duration) were recorded while the participants were describing the pathological mechanisms of sepsis. These post-intervention measurements varied in their timing related to the pre-intervention measurements due to variations in the timing during which the participants answered the questions. The simulation exercise was terminated after this point. Figure 2 demonstrates the anticipated cognitive load over time for simulation exercise 2.

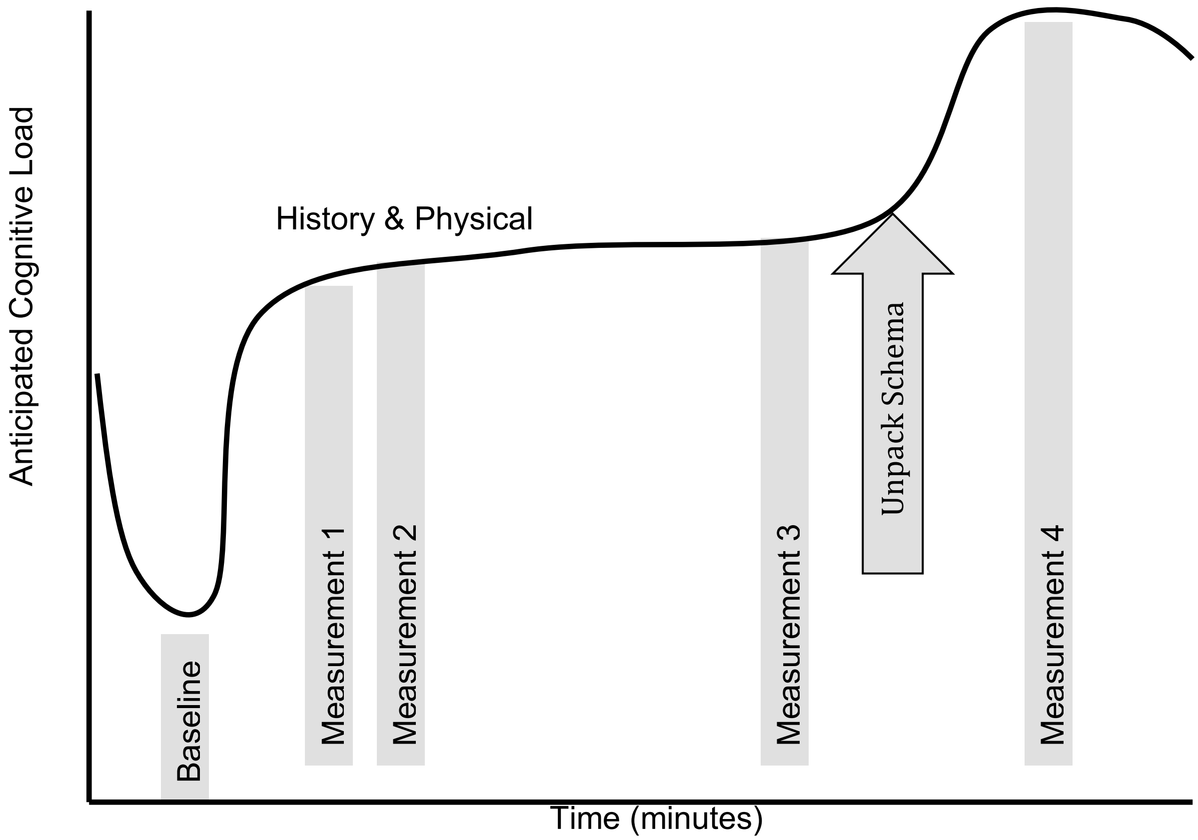

Figure 2. Schematic of cognitive load over time for simulation exercise 2.

Simulation Exercise 3: Algorithmic Approach. The simulated clinical scenario was one of iron toxicity with gastrointestinal haemorrhage (vomiting large amounts of blood), and ultimately cardiac arrest, in a middle aged man. There was very little ambiguity in the simulated clinical presentation, however the differential diagnosis (other things that could cause this presentation) was broad, and participants had to differentiate between several life threatening diseases simultaneously, all while providing lifesaving care to a critically ill patient. This theoretically resulted in a very high intrinsic cognitive load for the participants due to the significant complexity of the case. Participants obtained a clinical history and performed a physical examination of the mannequin. GSR and retrospective VAS measurements using video replay were obtained during this phase of the encounter. Data were collected for two 15 second intervals. These intervals were separated by 15 seconds. There was no experimental intervention between these first two measurements. The simulated patient's condition deteriorated into cardiac arrest. In this scenario, learners were coached to re-package elements into a schema, thereby decreasing the mental elements which must be handled by working memory. Specifically, participants were coached to ignore all the complex clinical information and differential diagnoses, and to treat the patient according to the Advanced Cardiac Life Support (ACLS) treatment algorithm for a pulseless electrical activity (PEA) cardiac arrest. This treatment algorithm emphasizes high quality cardiopulmonary resuscitation (CPR, chest compressions), minimizing pharmacological interventions, and quickly identifying and treating 14 reversible causes of cardiac arrest. While this seems complex to the lay observer, this approach focuses on algorithmic management, and all participants were well acquainted with this basic treatment algorithm. Post-intervention GSR and VAS measurements (15 seconds duration) were recorded while the participants were supervising the resuscitation of the mannequin while CPR was being performed and other treatment modalities were being considered to reverse the cause of the cardiac arrest. These post-intervention measurements varied in their timing related to the pre-intervention measurements due to variations in the timing during which the participants entered this algorithmic treatment approach. The simulation exercise was terminated after this point. Figure 3 demonstrates the anticipated cognitive load over time curve and interventions in exercise 3.

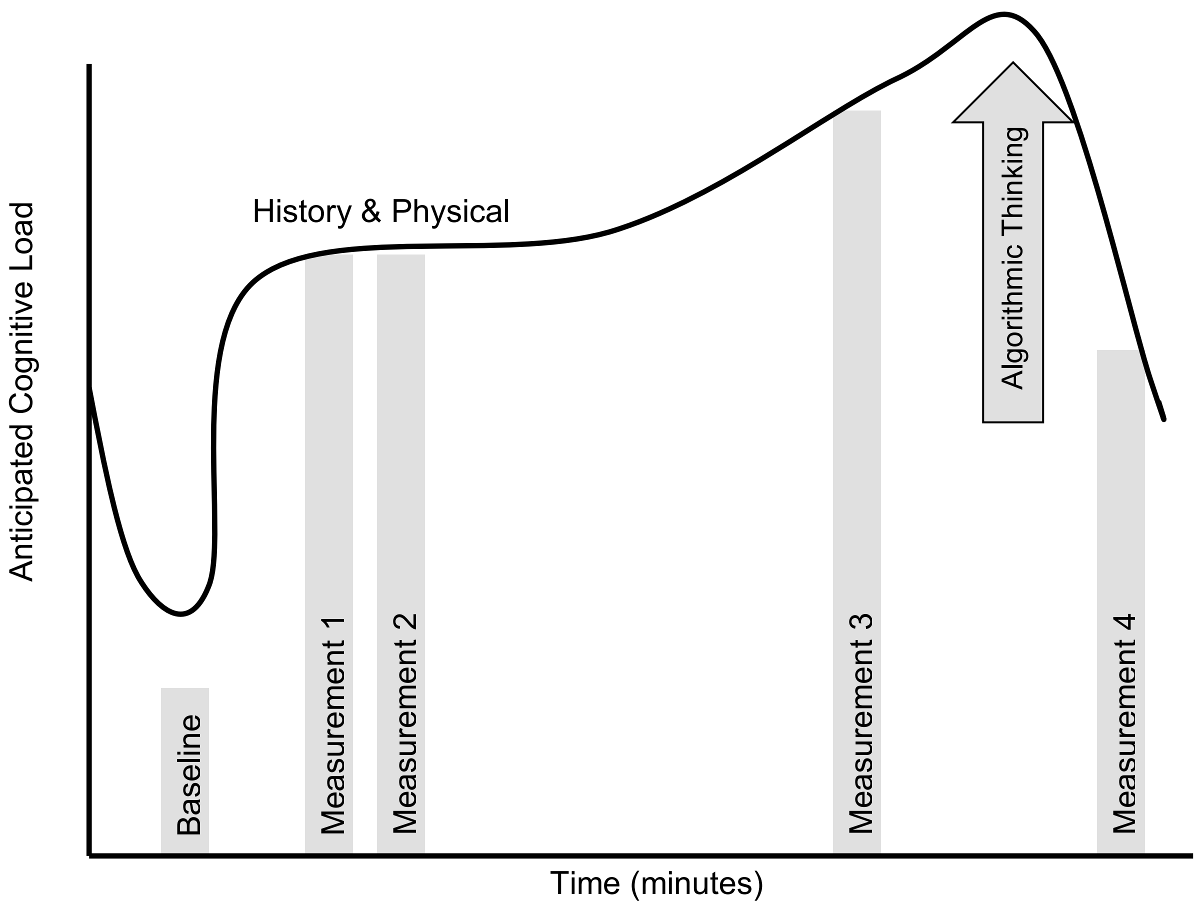

Figure 3. Schematic of cognitive load over time for simulation exercise 3.

Simulation Exercise 4: Worked Examples Effect. In this scenario, the participants were informed that they were assuming the role of a resident trauma physician, and that their attending physician was en-route to help, but that they were responsible for the initial care of an acutely ill trauma patient. The simulated patient had sustained multi-system trauma in an all terrain vehicle (ATV) accident.

The simulated patient had sustained massive brain injuries resulting in coma, a massive hemothorax (blood inside the thoracic cavity) as well as hemoperitoneum (blood in the abdominal cavity). GSR and retrospective VAS measurements using video replay were obtained while the participant was obtaining a clinical history and performing a physical examination. Data were collected for two 15 second intervals. These intervals were separated by 15 seconds. There was no experimental intervention between these first two measurements. After the participant identified all the injuries sustained by the simulated patient, a 15 second measurement of GSR and VAS was performed. A confederate actor entered the room and identified himself or herself as the attending trauma physician after these data were collected. The participant provided a brief synopsis of the patient encounter and was instructed to stand at the foot of the bed and watch (without any active participation) as the attending physician provided all subsequent care to the patient. The confederate attending physician did not provide any teaching or instruction while he, or she, cared for the patient. The participant simply watched the care provided. A final 15 second measurement of GSR and VAS was performed near the completion of the resuscitation by the attending physician, and the simulation exercise was terminated. Figure 4 depicts the anticipated cognitive load over time as well as the interventions during exercise 4.

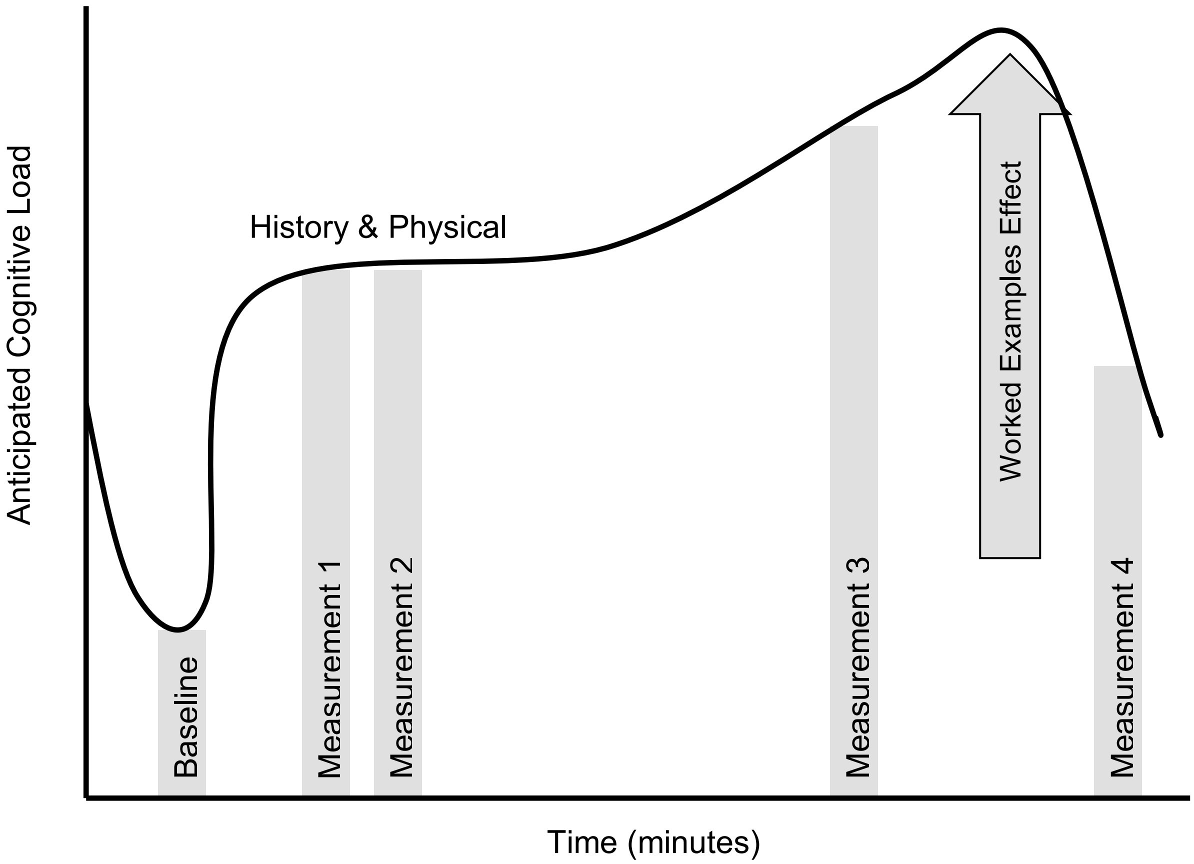

Figure 4. Schematic of cognitive load over time for simulation exercise 4.

CHAPTER FOUR. Results

Twenty-one participants took part in the study (13 male, 8 female). Two thirds of the participants (n=14) were in their first post graduate year of residency training, while the remaining third (n=7) were in their second year of residency training. Twenty participants received their undergraduate medical degrees from a Canadian medical school, and one graduated from medical school in the United Arab Emirates. The mean age of the participants was 27 years (SD=2.1 years). Table 1 shows the residency programs in which the participants are registered and prior simulation experience. Participants had a mean of 33.6 hours (SD 20.4 hours) of simulation experience prior to participating in the study. Second year residents had more simulation experience (50.0 hours vs 25.3 hours, $t(19)=-3.15$, $p=0.005$, $\eta_p^2=0.34$) than first year residents. Emergency medicine residents had more prior simulation experience (46.5 vs 28.4 hours), but this difference did not reach statistical significance, $t(19)=-1.97$, $p=0.063$, $\eta_p^2=0.20$.

Table 1.

Postgraduate medical training program and previous simulation experience

Postgraduate Training Program	N	Percent	Simulation Experience (hours +/- 95%CI)
Emergency Medicine	6	28.6	46.5 (18.1-74.9)
Family Medicine	9	42.9	30.3 (17.3-43.2)
Anesthesiology	2	9.5	32.5 (-189.9-254.9)
Internal Medicine	4	19.0	22.0 (12.9-31.1)
Total	21	100.0	

The Speilberger Trait anxiety scale measures one's baseline level of anxiety. The scale consists of 20 questions using a 4-point Likert scale. The lowest possible score is 20, and the highest possible score is 80. Trait anxiety increases proportional to the total score. Our participants had a mean Speilberger Trait Anxiety Score of 35.48 (SD=6.8). The internal consistency of the Speilberger trait anxiety score in our cohort was high (Cronbach's Alpha 0.87).

Data were analyzed using repeated measures ANCOVA and paired samples *t*-tests. When Mauchly's test of sphericity rejected the null hypothesis of sphericity, indicating that variances were not equal, corrections to within-subjects effects were calculated using the Greenhouse-Geisser method. Sphericity was not calculated when fewer than 3 data points were available for analysis. The Bonferroni method was used to correct for multiple comparisons. Significant interactions were graphed to ease interpretation.

Visual Spatial Memory Task

We undertook an analysis of visual spatial memory task responses in order to identify cognitive load data that best reflected the cognitive effort expended by participants during the visual spatial memory test. We were interested in the cognitive load involved in performing each task, analyzing the (a) data (all items) did not seem appropriate; for example, if participants disengaged from a task, they could show very low cognitive load. On the other hand, requiring completely correct responses ran the risk of eliminating too many responses for analysis. We chose to analyze the $\geq$75% responses as a compromise; to recall $\geq$75% of the elements in the more difficult items demonstrates considerable engagement, while allowing for a small number of errors. The descriptive statistics for VAS and GSR for each item of the visual spatial memory test are shown in Tables 2 and 3, (a) for all items regardless of whether the participant was correct, (b) when the answer was completely correct, and (c) when the participant achieved ≥75% of the correct responses.

Complete VAS data were available for all 21 participants. Failure of GSR device data transmission resulted in the loss of GSR data for 2 participants, leaving compete data available for analysis in 19.

The mean VAS and GSR measurements both increased as the number of items to recall increased. The VAS and GSR values correlated well with the number of task elements. The Pearson correlation coefficients for all responses were 0.97 and 0.95 for VAS and GSR data, respectively. When only correct responses were considered, the Pearson correlation coefficients for VAS and GSR data with the number of task elements were 0.98 and 0.35, respectively. When responses where participants achieved correct recall of ≥75% of dot positions were considered, the Pearson correlation coefficient for VAS and GSR data with the number of task elements were 0.98 and 0.88, respectively.

Table 2.

Descriptive statistics for visual spatial memory test visual analogue scale (VAS) measures of cognitive load.

Visual Spatial Task Item	Number of elements	n	Mean VAS (mm)	SD
		All Responses		
1	2	21	3.96	5.12
2	3	21	5.58	5.81
3	4	21	12.46	8.49
4	5	21	25.45	11.51
5	6	21	30.47	15.71
6	7	21	41.81	19.60
7	8	21	59.42	17.72
8	9	21	62.72	17.77
9	10	21	72.97	16.31
10	11	20	64.89	28.48
11	12	21	70.50	25.65
		Perfect Responses		
1	2	21	3.96	5.12
2	3	21	5.58	5.81
3	4	21	12.46	8.49
4	5	18	24.11	11.41
5	6	20	28.88	14.27
6	7	15	43.62	21.22
7	8	12	54.90	17.03
8	9	11	53.97	15.38
9	10	8	61.12	14.84
10	11	0		
11	12	1	66.69	
		Responses ≥75% Correct		
1	2	21	3.96	5.12
2	3	21	5.58	5.81
3	4	21	12.46	8.49
4	5	20	24.51	10.94
5	6	21	30.47	15.71
6	7	19	40.60	20.12
7	8	18	56.86	17.64
8	9	16	58.98	17.51
9	10	12	64.62	13.82
10	11	2	71.89	4.97
11	12	7	65.54	26.50

Table 3.

Visual spatial memory test Galvanic Skin Response (GSR) measures of cognitive load.

Visual Spatial Task Item	Number of Items to Recall	n	Mean GSR $x10^5$ (Arb)	SD
All Responses				
1	2	19	12.94	1.35
2	3	19	13.12	1.26
3	4	19	13.21	1.21
4	5	19	13.33	1.15
5	6	19	13.42	1.15
6	7	19	13.51	1.09
7	8	19	13.51	1.16
8	9	19	13.58	1.14
9	10	19	13.60	1.16
10	11	19	13.61	1.16
11	12	19	13.68	1.18
Perfect Responses				
1	2	19	12.94	1.35
2	3	19	13.12	1.26
3	4	19	13.21	1.21
4	5	16	13.49	1.00
5	6	18	13.34	1.13
6	7	14	13.62	1.09
7	8	10	13.54	1.24
8	9	10	13.33	1.11
9	10	7	13.03	1.21
10	11	0		
11	12	0		
Responses $\geq$75% Correct				
1	2	19	12.94	1.35
2	3	19	13.12	1.26
3	4	19	13.21	1.21
4	5	18	13.42	1.12
5	6	19	13.42	1.15
6	7	18	13.44	1.09
7	8	16	13.43	1.21
8	9	14	13.66	1.12
9	10	10	13.32	1.10
10	11	2	13.69	1.07
11	12	6	13.88	9.14

Table 4 summarizes the correlations between the GSR and VAS data for each subject's visual spatial memory task. Overall, the visual analogue scale (VAS) self-reports of total cognitive load were highly correlated with the galvanic skin response (GSR) data. When VAS and GSR data were filtered to only include (a) data from totally correct responses, or (b) responses in which participants correctly recalled $\geq$75% of the dot positions, the correlations between VAS and GSR scores increased. The mean correlations (across participants) ranged from 0.76 to 0.82, indicating a high degree of consistency between the two measures.

Cronbach's alpha was 0.86 for the VAS data and 0.997 for the GSR data during the visual spatial task when all responses were considered. The split-half reliability with Spearman-Brown correction was 0.69 for both VAS and GSR measurements when only perfect responses were considered. The split-half reliability with Spearman-Brown correction was 0.81 and 0.88 for VAS and GSR, respectively, when only responses with $\geq$75% correct recall were considered.

Table 4.

Pearson correlation coefficients between VAS and GSR data for visual spatial memory task when all responses are considered, when only perfect responses are considered, and when responses where participants correctly recalled ≥75% of dot positions.

Participant	All Responses	Only Perfect Responses	Responses where ≥75% Correct
1	0.90	0.91	0.94
2	0.80	0.78	0.74
3	-0.33	-0.60	0.41
4	0.68	0.81	0.81
5	0.91	0.86	0.86
6	0.82	0.87	0.84
7	0.94	0.89	0.90
8	0.94	0.87	0.89
9	0.93	0.92	0.91
10	0.92	0.89	0.92
11	0.95	0.91	0.95
12	0.96	0.98	0.98
13	0.80	0.80	0.81
14	0.55	0.63	0.62
15	0.83	0.94	0.92
16	*	*	*
17	0.46	0.55	0.53
18	0.98	0.96	0.98
19	*	*	*
20	0.36	0.52	0.58
21	0.98	0.99	0.99
Mean	0.76	0.76	0.82

* Data not available due to equipment malfunction

Given the superior reliability and correlation coefficients between VAS and GSR data, and between VAS or GSR data to the number of items for recall, we chose to perform our ANCOVA analysis on data in which participants correctly recalled $\geq$75% of the dot positions correctly during the visual spatial memory task. Sample GSR data of a participant is presented in Figure B1 (p. 84). Repeated measures analyses of covariance were carried out, with the repeated measure being the separate memory items (8 items, in order to preserve enough residual degrees of freedom for analysis), and the covariates being, in separate analyses, trait anxiety, age, and gender. In all of the analyses, including both VAS and GSR data, Mauchly's *W* was significant (all *W*'s=0.00, p<0.001), indicating that the null hypothesis of sphericity was rejected, and that variances were not equal. Corrections to within-subjects effects were calculated using the Greenhouse-Geisser method. When considering VAS with no covariates, there was a difference across items, Wilks' Λ = 0.06, F (2.72, 35.32) = 58.34, p < 0.001. The equivalent analysis of GSR showed Wilks' Λ = 0.11, F (1.62, 19.49) = 19.13, p<0.001.

In every analysis except that with age as a covariate for GSR, neither the main effect of the covariate nor the interaction of the covariate and the repeated measure was significant (all F's < 0.96, all p's > 0.35), and inclusion of the covariate did not change the interpretation of the main effect (Table 5).

As seen in Table 5, participant age did have a significant effect on GSR, F(1, 11)=8.08, p=0.02, η_p^2=0.42. GSR increased with age across all 8 memory test items. However, this effect did not change the overall strength of the item effect on

GSR, $F(1.67, 18.39)=21.08$, $p<0.001$, $\eta_p^2=0.66$.

We compared each pairwise (i.e. 1 vs. 2, 2 vs. 3, etc.) mean for both VAS and GSR data for the first 8 memory test items using paired samples *t*-tests (means are shown in Tables 2 and 3). All differences were statistically significant except the mean VAS differences between Stages 4 and 5, and Stages 7 and 8. Mean GSR differences were non-significant between Stages 6 and 7, and Stages 7 and 8.

Overall, in this standardized, static visual spatial memory test, both GSR and VAS were able to differentiate differing levels of cognitive load. VAS was more sensitive at detecting smaller differences between tasks, given that all of the VAS effects were stronger than the GSR effects in this visual spatial memory task.

Table 5.

Repeated measures analysis of covariance of visual spatial memory task considering responses where participants recalled >75% correctly

	Item					Item*Covariate				Covariate Effect			
Covariate	Wilks' Λ	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*
Visual Analogue Scale (mm)													
None	0.06	2.72, 35.32	58.34	0.82	<0.001								
Trait Anxiety	0.06	2.63, 31.53	55.51	0.82	<0.001	2.63, 31.53	0.37	0.03	0.75	1, 12	0.27	0.02	0.61
Age	0.06	2.70, 32.43	55.41	0.82	<0.001	2.70, 32.34	0.35	0.03	0.77	1, 12	0.07	0.01	0.80
Gender	0.06	2.81, 33.68	58.11	0.83	<0.001	2.81, 33.68	0.95	0.07	0.42	1, 12	0.96	0.07	0.35
Galvanic Skin Response (Arb)													
None	0.11	1.624, 19.49	19.13	0.61	<0.001								
Trait Anxiety	0.10	1.62, 17.83	17.72	0.62	<0.001	1.621, 17.832	0.12	0.01	0.85	1, 11	0.17	0.02	0.69
Age	0.01	1.67, 18.39	21.08	0.66	<0.001	1.672, 18.394	2.22	0.17	0.13	1, 11	8.08	0.42	0.02
Gender	0.11	1.62, 17.83	18.14	0.62	<0.001	1.621, 17.834	0.38	0.03	0.64	1, 11	0.37	0.03	0.56

Simulation Scenarios

Following measurement of cognitive load in a static and controlled environment, participants took part in four simulation scenarios. Each scenario was designed with an intervention to either increase or decrease cognitive load. In each of the following simulation exercises, the participants had their cognitive load measured in a dynamic environment while they interacted with other people, as well as a simulated patient. Each scenario is summarized below, and results of VAS and GSR measurements are reported using repeated measures ANCOVA.

Simulation Exercise 1: Increasing Element interactivity

In this simulated clinical experience, VAS and GSR data were both collected at 4 time points. Baseline GSR (but not VAS) was also recorded as the participant sat quietly in the simulation laboratory without any other activity occurring. The first two measurements were taken at the beginning of the scenario, when the participants were obtaining a clinical history from the mannequin. The second two measurements were taken after the participants had obtained a clinical history, performed a physical exam, interpreted a chest X-ray, and interpreted an electrocardiogram. The mean values of the two measurements at the beginning of the scenario were compared to the mean values of the two measurements at the end of the scenario for both GSR and retrospective VAS. Figure 1 (Chapter 3, p. 29) depicts the hypothetical cognitive load response curve during the simulation exercise, and describes the participants' experiences and the time points when measurements of GSR and VAS were taken. Figure B2 (Appendix B, p. 85) represents actual GSR data for a participant during the first simulation exercise.

Complete VAS and GSR data were available for analysis for all participants. Repeated measures analyses of covariance were carried out, with the repeated measure (Time) being the VAS or GSR data at the two time points described above, and the covariates being, in separate analyses, trait anxiety, age, gender, baseline GSR (only for GSR data) and simulation experience. Age had a covariate effect on VAS, where VAS increased with age, $F(1, 19)=2.06$, $p=0.01$, $\eta_p^2=0.30$, and baseline GSR demonstrated a covariate effect on GSR data where GSR was proportional to baseline GSR, $F(1,19)=167.11$, $p<0.001$, $\eta_p^2=0.90$. Despite these effects, the overall interpretation of our results was unaffected. In every other analysis, neither the main effect of the covariate nor the interaction of the covariate and the repeated measure was significant (all F's < 4.23, all p's > 0.05), and did not change the interpretation of the repeated measures main effect. The following results are from the analysis with no covariates.

Mean VAS was significantly greater at the end of the scenario than at the beginning. Mean VAS increased from 44.78mm at the beginning of the scenario, to 65.54 mm at the end of the scenario, $F(1, 20)=42.69$, $p<0.001$, $\eta_p^2=0.68$. Mean GSR also increased between the beginning and end of this scenario, although this difference did not reach statistical significance. Mean GSR was 5.53×10^5 Arb at the outset, and 5.54×10^5 Arb) at the end, $F(1, 20)=0.47$, p=0.50, $\eta_p^2=0.02$. Table 6 summarizes the descriptive statistics for exercise 1.

The intervention in this exercise was designed to increase the number of elements handled in the participants' working memory and was expected to increase total cognitive load. VAS measurement was able to detect this difference, while GSR was not.

Table 6.

Descriptive statistics for VAS and GSR data during simulation exercise 1.

Measure	Mean	Standard Deviation
VAS Pre-Intervention (mm)	44.78	14.62
VAS Post-Intervention (mm)	65.54	15.83
GSR Baseline (Arb x 10^5)	5.43	0.46
GSR Pre-Intervention (Arb x 10^5)	5.53	0.43
GSR Post-Intervention (Arb x 10^5)	5.54	0.42

Table 7.

Repeated measures analysis of covariance for simulation exercise 1.

	Time					Time*Covariate				Covariate Effect			
Covariate	Wilks' Λ	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*
Visual Analogue Scale (mm)													
None	0.32	1, 20	42.69	0.68	<0.001								
Trait Anxiety	0.32	1, 19	40.58	0.68	<0.001	1, 19	0.003	0.00	0.96	1, 19	0.01	0.001	0.91
Age	0.30	1, 19	44.94	0.70	<0.001	1, 19	2.06	0.10	0.17	1, 19	8.27	0.30	0.01
Gender	0.30	1, 19	45.24	0.70	<0.001	1, 19	2.20	0.10	0.16	1, 19	0.02	0.001	0.89
Simulation Experience	0.32	1, 19	40.88	0.68	<0.001	1, 19	0.16	0.01	0.70	1, 19	0.12	0.01	0.73
Galvanic Skin Response (Arb)													
None	0.98	1, 20	0.47	0.02	0.50								
Trait Anxiety	0.98	1, 19	0.45	0.02	0.51	1, 19	0.12	0.01	0.74	1, 19	0.02	0.001	0.88
Age	0.97	1, 19	0.54	0.03	0.47	1, 19	4.08	0.18	0.06	1, 19	4.23	0.18	0.05
Gender	0.98	1, 19	0.47	0.02	0.50	1, 19	1.04	0.05	0.32	1, 19	0.18	0.01	0.68
Baseline GSR	0.97	1, 19	0.52	0.03	0.48	1, 19	2.88	0.13	0.11	1, 19	167.11	0.90	<0.001
Simulation Experience	0.97	1, 19	0.51	0.03	0.48	1, 19	2.70	0.12	0.12	1, 19	1.18	0.09	0.19

Simulation Exercise 2: Unpack Schema

In this exercise, participants resuscitated a patient who ultimately suffered a cardiac arrest due to pneumonia and sepsis. Baseline GSR was recorded as the participant sat quietly in the simulation laboratory without any other activity occurring. GSR and retrospective VAS data were both collected at 4 points. The first two measurements were taken while the participants obtained a history from the simulated patient. Measurements of VAS and GSR were also recorded before, and after, our intervention. The intervention was designed to have the participant "unpack" their schema of sepsis, by having a confederate medical student ask them questions about the basic pathophysiology of sepsis leading to cardiac arrest, and thus to increase cognitive load. The GSR and VAS values are compared before and after our intervention. Figure 2 (Chapter 3, p. 31) depicts the anticipated cognitive load response curve during the simulation exercise, and describes the participants' experiences and the time points when measurements of GSR and VAS were taken. Figure B3 (Appendix B, p. 86) depicts actual participant GSR data during the second simulation exercise. Complete VAS and GSR data were available for analysis for all participants.

Repeated measures analyses of covariance were carried out, with the repeated measure (Time) being the VAS and GSR data before and after our intervention, and the covariates being, in separate analyses, trait anxiety, age, gender, baseline GSR (only for GSR data) and simulation experience.

Age had a significant covariate effect on VAS, where VAS increased with age, $F(1, 19)=6.78$, $p=0.02$, $\eta_p^2 = 0.26$. Both the Baseline GSR, $F(1,19)=230.01$,

p<0.001, η_p^2 =0.92, and Time x Simulation Experience, F(1,19)=4.84, p=0.04, η_p^2 = 0.20, were significant. As seen in Figure 5, the interaction shows that participants with greater simulation experience have lower GSR scores. This relationship between simulation experience and GSR is weaker after our intervention (slope is less steep), and much of the effect seems to be due to two participants with widely differing pre-and post-intervention GSR scores. When these two participants are excluded from analysis, the interaction of simulation experience with time becomes non-significant (F(1,17)=0.321, p=0.579, η_p^2 =0.019), and this change does not affect our overall interpretation of our results. The change in the relationship before and after intervention is small, and unlikely to be an important influence on our major findings. In every other analysis, neither the main effect of the covariate nor the interaction of the covariate and the repeated measure was significant (all F's < 3.26, all p's > 0.09), and did not change the interpretation of the main effect. Results of our repeated measures ANCOVA for exercise 2 are presented in Table 9.

Retrospective VAS scores were significantly higher after intervention (77.67 mm) compared to before the intervention (52.87 mm), F(1, 20)=16.21, p=0.001, η_p^2 = 0.45. The mean GSR value was 5.52×10^5 Arb before the intervention and 5.49×10^5 Arb after the intervention, F(1, 20)=0.76, p=0.39, η_p^2 = 0.04. Table 8 summarizes the descriptive statistics for exercise 2. Similar to the first exercise, VAS measurement was able to detect differences in cognitive load, while GSR measurement was not.

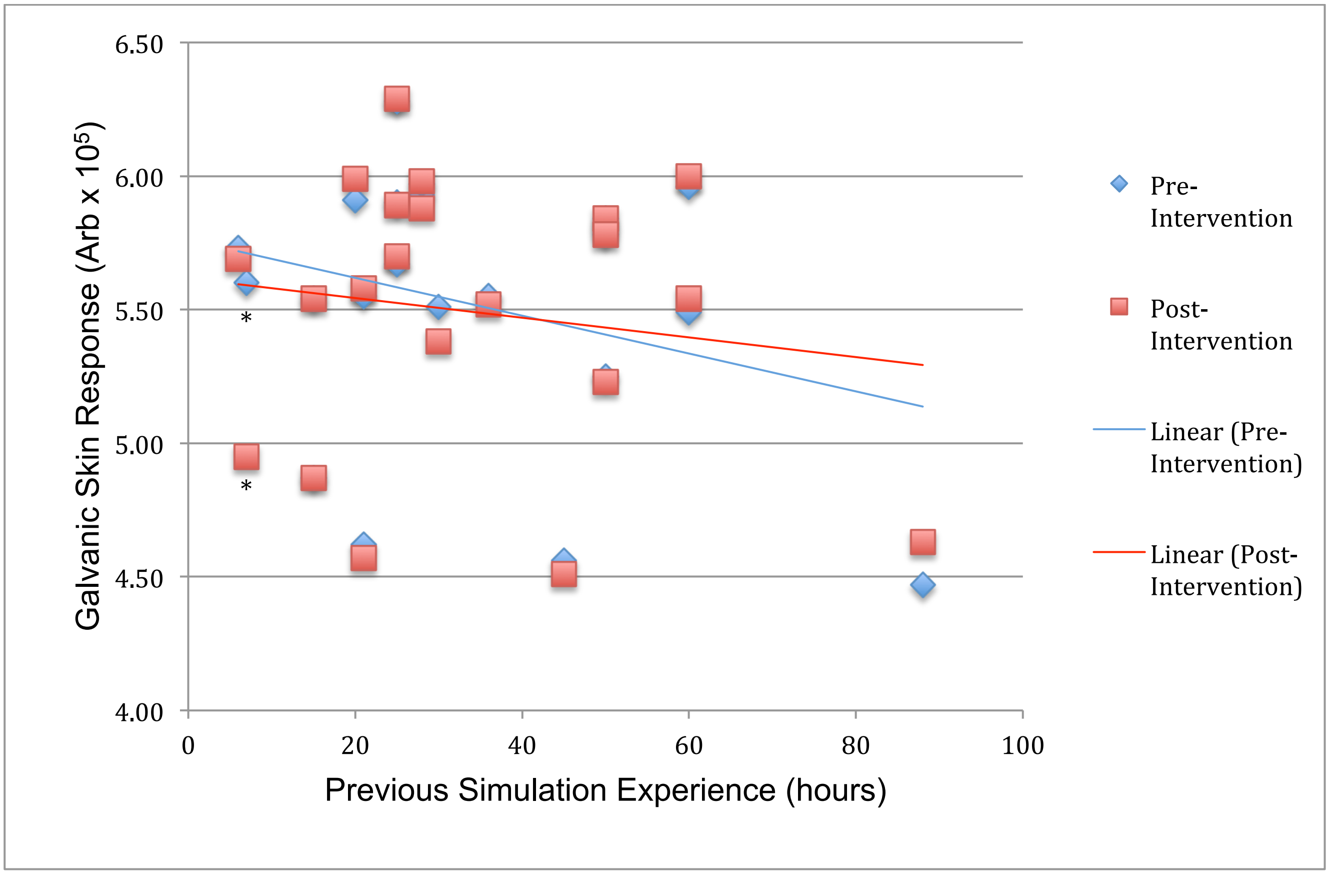

Figure 5. Interaction between previous simulation experience and Time during exercise 2. * Identifies the individual participant with widely different GSR results pre- and post-intervention.

Table 8.

Descriptive statistics for VAS and GSR data during simulation exercise 2.

Measure	Mean	Standard Deviation
VAS Pre-Intervention (mm)	52.87	22.43
VAS Post-Intervention (mm)	77.67	18.26
GSR Baseline (Arb x 10^5)	5.44	0.52
GSR Pre-Intervention (Arb x 10^5)	5.52	0.50
GSR Post-Intervention (Arb x 10^5)	5.49	0.51

Table 9.

Repeated measures analysis of covariance for exercise 2

	Time					Time*Covariate				Covariate Effect			
Covariate	Wilks' Λ	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*
Visual Analogue Scale (mm)													
None	0.55	1, 20	16.21	0.45	0.001								
Trait Anxiety	0.55	1, 19	15.81	0.45	0.001	1, 19	0.51	0.03	0.49	1, 19	1.35	0.07	0.26
Age	0.53	1, 19	16.64	0.47	0.001	1, 19	1.53	0.07	0.23	1, 19	6.78	0.26	0.02
Gender	0.55	1, 19	15.48	0.45	0.001	1, 19	0.11	0.01	0.75	1, 19	1.72	0.08	0.20
Simulation Experience	0.54	1, 19	16.53	0.47	0.001	1, 19	1.40	0.07	0.25	1, 19	0.09	0.01	0.77
Galvanic Skin Response (Arb)													
None	0.96	1, 20	0.76	0.04	0.39								
Trait Anxiety	0.96	1, 19	0.84	0.04	0.37	1, 19	3.17	0.14	0.09	1, 19	0.000	0.000	0.99
Age	0.96	1, 19	0.72	0.04	0.41	1, 19	0.04	0.002	0.85	1, 19	3.26	0.15	0.09
Gender	0.96	1, 19	0.80	0.04	0.38	1, 19	2.20	0.10	0.16	1, 19	0.87	0.04	0.36
Baseline GSR	0.96	1, 19	0.72	0.04	0.41	1, 19	0.08	0.004	0.79	1, 19	230.01	0.92	<0.001
Simulation Experience	0.96	1, 19	0.90	0.05	0.35	1, 19	4.84	0.20	0.04	1, 19	0.95	0.05	0.34

Simulation Exercise 3: Algorithmic Approach

In this third simulated clinical encounter, the participants care for a very complex patient who has overdosed on iron tablets. Ultimately, the patient suffers a cardiac arrest, and the participants are coached by a confederate actor to approach the resuscitation using a standardized algorithmic approach to cardiac arrest. Baseline GSR was recorded as the participant sat quietly in the simulation laboratory without any other activity occurring. GSR and retrospective VAS data were both collected at 4 points. The first two measurements were taken while the participants obtained a history from the simulated patient. Measurements of VAS and GSR were also recorded before and after our intervention, which was anticipated to decrease cognitive load. The GSR and VAS values are compared before and after our intervention. Figure 3 (Chapter 3, p. 33) depicts the anticipated cognitive load response curve during the simulation exercise, and describes the participants' experiences and the time points when measurements of GSR and VAS were taken. Figure B4 (Appendix B, p. 87) depicts actual participant GSR data during the third simulation exercise. GSR data was not available for analysis (due to data transmission failure of the GSR device) in two participants, leaving 19 participants for analysis.

Repeated measures analyses of covariance were carried out, with the repeated measure (Time) being the VAS and GSR data before and after our intervention, and the covariates being, in separate analyses, trait anxiety, age, gender, baseline GSR (only for GSR data) and simulation experience.

Age had a significant covariate effect on GSR, where GSR increased with age, $F(1,17)=4.93$, $p=0.04$, $\eta_p^2=0.23$). Additionally, both the interaction of baseline GSR and Time ($F(1,17)=7.06$, $p=0.02$, $\eta_p^2=0.29$) and the covariate baseline GSR ($F(1,17)=459.44$, $p<0.001$, $\eta_p^2=0.96$) had significant effects on GSR, where GSR was proportional to baseline GSR. None of these effects change our interpretation of our main effects (Table 11). Figure 6 depicts the relationship of baseline GSR to both pre- and post-intervention GSR. While the interaction of baseline GSR and Time was statistically significant, the slopes of both curves (baseline GSR predicting pre-intervention GSR, and baseline GSR predicting post-intervention GSR) are nearly identical, indicating that this interaction is not important. This interaction does not alter the interpretation of our main results, as seen in Table 11.

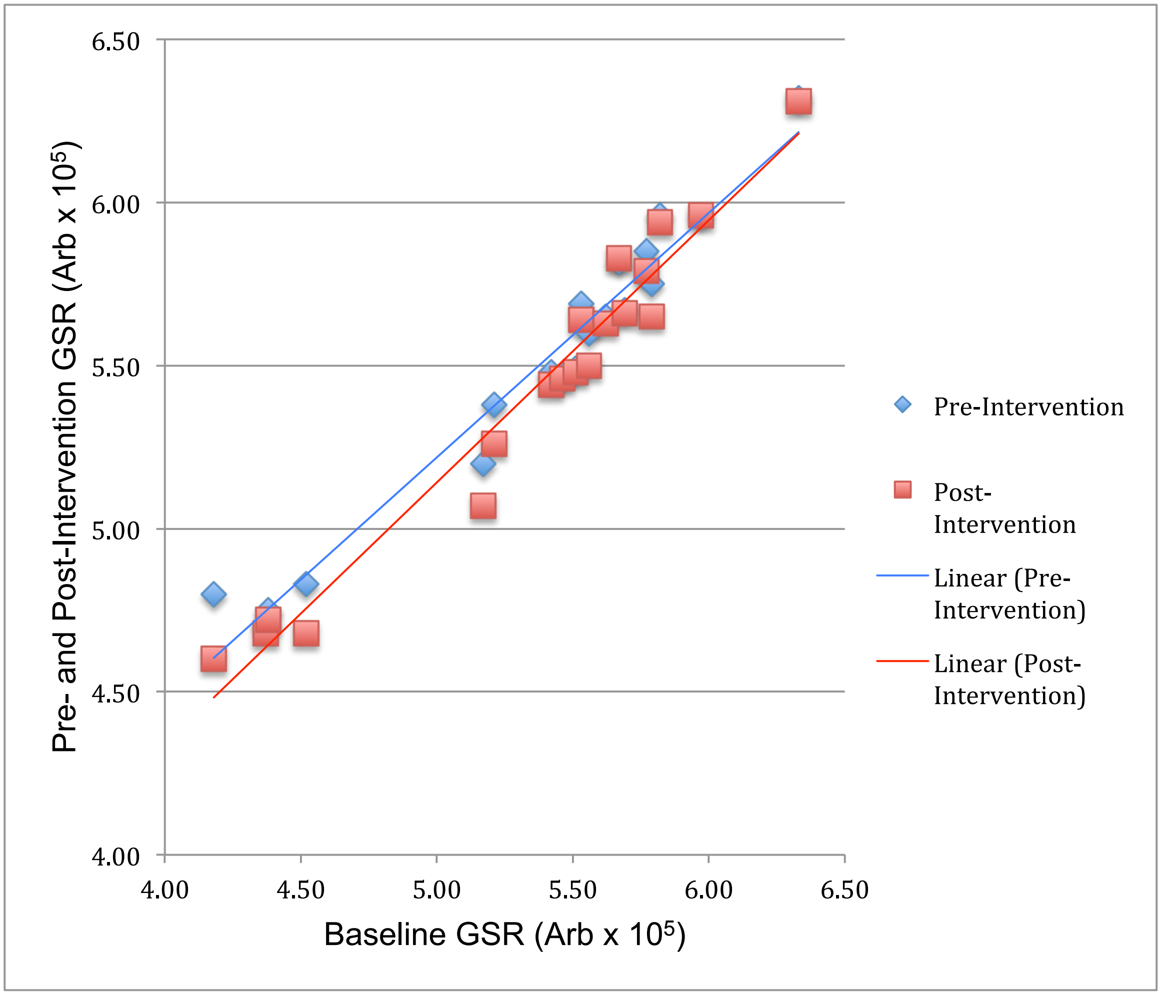

Figure 6. Interaction of baseline GSR data with Time in simulation exercise 3.

In every other analysis, neither the main effect of the covariate nor the interaction of the covariate and the repeated measure was significant (all *F*'s < 1.79, all *p*'s > 0.20), and did not change the interpretation of the main effect (Table 11).

Self-reported cognitive load using the retroactive VAS decreased after our intervention. The mean VAS value was 63.65 mm before the intervention, and 55.17 mm after the intervention. This difference approached, but did not achieve, statistical significance, $F(1, 19)=4.01$, $p=0.06$, $\eta_p^2=0.17$. Mean GSR scores also decreased after intervention, from 5.49×10^5 Arb to 5.43×10^5 Arb. This difference did reach statistical significance, $F(1, 18)=17.95$, $p<0.001$, $\eta_p^2 = 0.50$. Table 10 summarizes the descriptive statistics for Exercise 3.

In this exercise, we expected our intervention to decrease measured cognitive load. While VAS measurement trended towards statistical significance, the decrease in VAS measurement was not significant. In this exercise, GSR measurement was able to identify a statistically significant decrease in cognitive load.

Table 10.

Descriptive statistics for VAS and GSR data during simulation exercise 3.

Measure	Mean	Standard Deviation
VAS Pre-Intervention (mm)	63.65	19.48
VAS Post-Intervention (mm)	55.17	25.16
GSR Baseline (Arb x 10^5)	5.37	0.59
GSR Pre-Intervention (Arb x 10^5)	5.49	0.45
GSR Post-Intervention (Arb x 10^5)	5.43	0.48

Table 11.

Repeated measures analysis of covariance for exercise 3

	Time					Time*Covariate				Covariate Effect			
Covariate	Wilks' Λ	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*
Visual Analogue Scale (mm)													
None	0.83	1, 19	4.01	0.17	0.06								
Trait Anxiety	0.82	1, 18	3.84	0.18	0.07	1, 18	0.23	0.01	0.64	1, 18	0.07	0.004	0.80
Age	0.82	1, 18	3.90	0.18	0.06	1, 18	0.50	0.03	0.49	1, 18	1.79	0.09	0.20
Gender	0.83	1, 18	3.81	0.18	0.07	1, 18	0.06	0.003	0.81	1, 18	0.59	0.03	0.45
Simulation Experience	0.82	1, 18	3.89	0.18	0.06	1, 18	0.46	0.03	0.51	1, 18	0.02	0.001	0.90
Galvanic Skin Response (Arb)													
None	0.50	1, 18	17.95	0.50	<0.001								
Trait Anxiety	0.50	1, 17	16.97	0.50	0.001	1, 17	0.02	0.001	0.89	1, 17	0.06	0.003	0.82
Age	0.48	1, 17	18.50	0.52	<0.001	1, 17	1.56	0.08	0.23	1, 17	4.93	0.23	0.04
Gender	0.49	1, 17	17.69	0.51	0.001	1, 17	0.75	0.04	0.40	1, 17	0.75	0.04	0.40
Baseline GSR	0.42	1, 17	23.99	0.59	<0.001	1, 17	7.06	0.29	0.02	1, 17	459.44	0.96	<0.001
Simulation Experience	0.49	1, 17	17.79	0.51	0.001	1, 17	0.84	0.05	0.37	1, 17	1.09	0.06	0.31

Simulation Exercise 4: Worked Examples Effect

In this scenario, the participants were tasked with the resuscitation of a critically ill trauma patient. Baseline GSR was recorded as the participant sat quietly in the simulation laboratory without any other activity occurring. GSR and retrospective VAS data were both collected at 4 points. The first two measurements were taken while the participants obtained a history from the simulated patient. Measurements of VAS and GSR were also recorded before and after our intervention. Our intervention was to introduce the worked examples effect and decrease cognitive load, by having a confederate attending trauma specialist enter the room and take over the resuscitation of the patient. Following a brief case summary and hand-over of care to the confederate trauma team leader, the participants were instructed to observe the resuscitation, but not to participate actively. VAS and GSR values were compared before and after our intervention. Figure 4 (Chapter 3, p. 35) depicts the anticipated cognitive load response curve during the simulation exercise, and describes the participants' experiences and the time points when measurements of GSR and VAS were taken. Figure B5 (Appendix B, p. 88) depicts actual GSR data for a participant during the final simulation exercise. GSR data was not available for analysis (due to data transmission failure of the GSR device) in two participants, leaving 19 participants for analysis.

Repeated measures analyses of covariance were carried out, with the repeated measure (Time) being the VAS and GSR data before and after our intervention, and the covariates being, in separate analyses, trait anxiety, age, gender, baseline GSR (only for GSR data) and simulation experience.

Age had a significant covariate effect on GSR data, where GSR increased with age, $F(1,17)=7.00$, $p=0.02$, $\eta_p^2=0.29$. This effect did not alter our interpretation of our main results (Table 13). Additionally, both the Time x Baseline GSR ($F(1,17)=5.60$, $p=0.03$, $\eta_p^2=0.25$) and Baseline GSR ($F(1,17)=50.48$, $p<0.001$, $\eta_p^2 = 0.75$) effects were significant. Figure 7 depicts the interaction. While the interaction was statistically significant, the slopes of both curves (baseline GSR to pre-intervention GSR, and baseline GSR to post-intervention GSR) are nearly identical, indicating that this interaction is not important. This interaction does not alter the interpretation of our main results, as seen in Table 13.

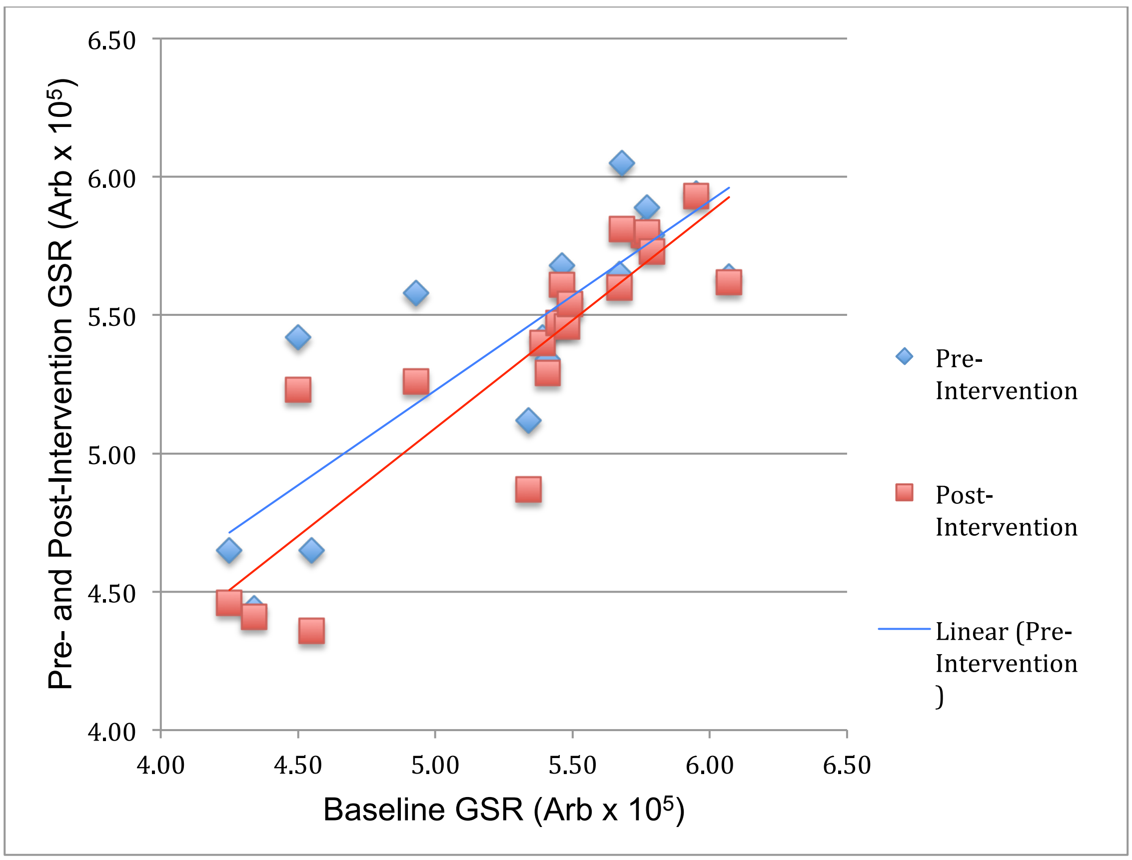

Figure 7. Interaction of Time and Baseline GSR in simulation exercise 4.

In every other analysis, neither the main effect of the covariate nor the interaction of the covariate and Time was significant (all *F’*s < 0.002, all *p’*s > 0.06), and did not change the interpretation of the main effect. Cognitive load decreased on both the VAS and GSR scales after intervention. VAS decreased from 65.85 mm to 27.87 mm. This difference reached statistical significance, $F(1, 19)=53.63$, $p<0.001$, $\eta_p^2=0.74$. Mean GSR decreased from 5.45×10^5 Arb to 5.35×10^5 Arb. This difference was also statistically significant, $F(1, 18)=19.32$, $p<0.001$, $\eta_p^2=0.52$. In this exercise, designed to decrease cognitive load, both VAS and GSR measurements were able to identify statistically significant decreases in total cognitive load.

Table 12.

Descriptive statistics for VAS and GSR data during simulation exercise 4.

Measure	Mean	Standard Deviation
VAS Pre-Intervention (mm)	65.85	14.81
VAS Post-Intervention (mm)	27.87	27.35
GSR Baseline (Arb x 10^5)	5.33	0.55
GSR Pre-Intervention (Arb x 10^5)	5.45	0.45
GSR Post-Intervention (Arb x 10^5)	5.35	0.49

Table 13.

Repeated measures analysis of covariance for exercise 4

	Time					Time*Covariate				Covariate Effect			
Covariate	Wilks' Λ	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*	*df*	*F*	η_p^2	*p*
Visual Analogue Scale (mm)													
None	0.26	1, 19	53.63	0.74	<0.001								
Trait Anxiety	0.24	1, 18	57.68	0.76	<0.001	1, 18	2.44	0.12	0.14	1, 18	0.85	0.05	0.370
Age	0.26	1, 18	50.90	0.74	<0.001	1, 18	0.04	0.002	0.85	1, 18	0.57	0.03	0.461
Gender	0.26	1, 18	52.00	0.74	<0.001	1, 18	0.42	0.02	0.52	1, 18	0.22	0.01	0.643
Simulation Experience	0.22	1, 18	62.29	0.78	<0.001	1, 18	4.07	0.18	0.06	1, 18	0.002	0.00	0.965
Galvanic Skin Response (Arb)													
None	0.48	1, 18	19.32	0.52	<0.001								
Trait Anxiety	0.47	1, 17	19.42	0.53	<0.001	1, 17	1.09	0.06	0.31	1, 17	0.15	0.01	0.700
Age	0.46	1, 17	20.24	0.54	<0.001	1, 17	1.85	0.10	0.19	1, 17	7.00	0.29	0.017
Gender	0.46	1, 17	19.69	0.54	<0.001	1, 17	1.34	0.07	0.26	1, 17	0.20	0.01	0.660
Baseline GSR	0.41	1, 17	24.26	0.59	<0.001	1, 17	5.60	0.25	0.03	1, 17	50.48	0.75	<0.001
Simulation Experience	0.47	1, 17	18.93	0.53	<0.001	1, 17	0.63	0.04	0.44	1, 17	1.80	0.10	0.197

CHAPTER FIVE: Discussion

Our goal with this research was to examine whether the manipulation and accurate measurement of cognitive load was possible in the dynamic environment of simulation-based medical education. To that end, we devised a series of five experiments, all framed within cognitive load theory. The standardized visual spatial memory task was designed as an objective validation tool for both our VAS and GSR measures of cognitive load. Both VAS and GSR measures of cognitive load were highly correlated with the number of recall items in this task, and both measures demonstrated high degrees of reliability. Furthermore, both measures were highly correlated with each other, again supporting the reliability of both VAS and GSR measures of cognitive load in this setting. The ultimate goal of this research was to apply these measures of cognitive load in an active and dynamic environment, so four simulation scenarios were created to evaluate VAS and GSR measures of cognitive load within the SBME environment. The first two scenarios were designed to increase cognitive load. The first was designed to add elements in to the participants' working memory (ECG interpretation, chest X-ray interpretation, changing clinical status of the mannequin), thereby increasing the intrinsic cognitive load. The second scenario was designed to increase the intrinsic cognitive load by having the participants unpack schemas during the scenario, in a sense forcing them to use multiple sub-components of their schema of sepsis instead of their single overall schema. The third and fourth scenarios were designed to decrease cognitive load. The third scenario was designed to have the participants approach a complex problem with high element interactivity using an algorithmic approach, with which the participants were well familiarized, thus decreasing the element interactivity and intrinsic cognitive load handled in their working memory. The fourth scenario was designed to evaluate the worked examples effect during SBME. This was designed to decrease their cognitive load.

While our participants were fairly homogenous with respect to their age, training, and experience, they are representative of the postgraduate medical trainee. While interpreting our results in reference to other populations should be done with caution, they should be considered valid for medical trainees.

Both VAS and GSR measures were able to identify the increased cognitive load associated with increased numbers of recall items in our visual spatial task. In the simulated scenarios, VAS was able to identify increases in cognitive load in the first two scenarios, while GSR was not. In the third scenario, GSR was able to identify decreases in cognitive load, while VAS measurement of cognitive load was not. Finally, in the fourth scenario, both measures of cognitive load were able to identify decreases in cognitive load (Table 14).

In a static environment (the standardized visual spatial memory task) both the VAS and GSR were able to demonstrate increasing levels of cognitive load. The effect sizes for VAS were consistently greater than for GSR in the visual spatial task (Table 5).

Table 14.

Effects of instructional interventions on total cognitive load as measured by retrospective VAS scales and GSR

Task	Intervention	Hypothesized Effect on Cognitive Load	Observed Effect on VAS Measurement	Observed Effect on GSR Measurement
Visual Spatial Memory Test	Increasing Elements	Increase	Increase	Increase
Exercise 1	Increasing Elements	Increase	Increase	NS
Exercise 2	Unpack Schema	Increase	Increase	NS
Exercise 3	Algorithmic Approach	Decrease	NS	Decrease
Exercise 4	Worked Examples Effect	Decrease	Decrease	Decrease

Note. NS= Non-significant change

During the first two dynamic tasks of simulated medical resuscitation, the effect sizes for VAS data were greater than those for GSR data (Tables 7 & 9). In the first exercise, task elements were added, and VAS measurement was able to identify statistically significant differences in cognitive load, while GSR was not. This lack of sensitivity was also seen in the second simulated exercise, where the unpacking of schemas should have increased cognitive load. Again, VAS measurement was able to identify this difference, while GSR data did not. Anecdotally, this simulated encounter seemed challenging for the majority of our participants, and this non-significant decrease in GSR data may have actually been due to cognitive overload and the withdrawal of cognitive resources as the participants became overwhelmed. In the absence of a mixed methods design including interviews of the participants, this remains speculative. Further study of the physiological response to cognitive overload is required using GSR.

The third simulation exercise was designed to decrease cognitive load, by using an algorithmic approach to resuscitation. Interestingly, GSR was able to identify statistically significant differences in cognitive load that VAS did not. In the fourth exercise, both GSR and VAS measures were able to identify decreases in cognitive load due to the introduction of the worked examples effect. Again, the VAS measures' effect sizes were larger than those of the GSR data in all realms of the ANCOVA. That said, GSR was able to identify statistically significant differences congruent with VAS measures when the differences were large enough. These findings suggest that our sample size may have been underpowered to identify differences that did exist. Given that we were unable to identify any previous studies of GSR measurement in SBME environments, we were unable to estimate an effect size for the calculation of a properly powered sample size.

Participant age demonstrated significant covariate effects on GSR data in the visual spatial memory test, Exercise 3, and Exercise 4. Age also had a significant covariate effect on VAS measurements in Exercise 1 and Exercise 2. In all cases, GSR and VAS measurements of cognitive load increased with age. None of these covariate effects changed our interpretation of the main effect. Our subjects had a mean age of 27 years (95%CI 26.4-28.3 years). It is unlikely that any physiological differences exist between a 26 year old and a 28 year old that would affect their vasomotor tone or sympathetic nervous system. There may be differences in cognitive processing as one ages, or gains more clinical experience as a medical trainee. These cognitive processing changes may alter the results of cognitive load measurement. Further inquiry along these lines is required to determine if age alone, or age related cognitive processes change measures of cognitive load.

Limitations

Limitations of this research include failing to control for vasoactive substance use, failing to include measures of “state” anxiety in addition to “trait” anxiety, and gender specific physiological differences. We also lacked a gold standard measure of cognitive load in a dynamic environment with which to compare our two measures. We compared two non-validated tools to each other. Currently, a gold standard measure of cognitive load in a dynamic environment does not exist. We were unable to compare measures of cognitive load in the simulations exercises to those of the simulation exercises. We reported total cognitive load, in both the simulation exercises and the standard memory test. Each measurement of cognitive load was 15 seconds long during the simulation exercises and 40 seconds long for each item in the memory test. While we could have reported mean cognitive load measures, and then compared the cognitive load experienced in the simulation exercise to that of the memory test, the two types of tasks varied significantly. One type of task was a dynamic and interactive experience (simulation exercises), while the other was a well-controlled, static, and quiet environment without interaction with other people (memory test). As a result of these differences, we elected not to directly compare the cognitive load measurements of the simulation exercises to the memory test.

Many substances affect the sympathetic nervous system, and the effects of catecholamines on end organs. Substances that stimulate the sympathetic nervous system, such as caffeine, produce vasoconstriction in the skin, and may decrease measured GSR values (Lader, 1969). While the use of other amphetamines, beta-blocking agents, or calcium channel blocking agents is less likely than the use of caffeine in our participant population, we do not know the exact incidence of their use, because we did not collect these data. Future studies using GSR should either quantify the serum concentrations of these substances, exclude those who use vasoactive substances, or more realistically, instruct their participants to abstain from their use prior to GSR measurement. Controlling for baseline GSR readings in our analysis would have controlled for the effects of these substances on our GSR measures of cognitive load.

Similar to pharmacologic confounders, hormonal differences between males and females, and hormonal differences based on the phase of the menstrual cycle in females can affect the sympathetic system's adrenergic response. In one study, men, and women in the luteal stage of their menstrual cycle, demonstrated a greater salivary cortisol response to a standardized social stress test than did women in the follicular phase of their menstrual cycle or those using oral contraceptive medications (Kirschbaum, Kudielka, Gaab, Schommer, & Hellhammer, 1999). Neither gender, nor gender's interaction with our factor of interest, had any influence on any of our effects. These differences may be more meaningful in a larger sample size, so future investigations should either control for oral contraceptive use, and the phase of menstrual cycle for female participants.

Finally, we controlled for trait anxiety level in our participants, but not their level of anxiety at the outset of the experiment. While our participants scored in the lower range of the Speilberger Trait Anxiety scale, we did not incorporate Speilberger's "State" anxiety scale. This state anxiety may have had a significant effect on our results that we did not appreciate. Healthy participants with higher trait anxiety levels produce attenuated physiological responses to stress. Individuals with higher trait anxiety levels have been shown to produce less epinephrine, norepinephrine, and cortisol compared to less anxious subjects. This results in the attenuation of the galvanic skin response to stress in individuals with higher levels of trait anxiety (Jezova et al., 2004). In this same study, all participants, regardless of their trait anxiety level showed increases in their state anxiety level after the application of a standardized psychological stress. The state anxiety level increased more in the higher trait anxiety group. Thus, participants with higher trait anxiety may have had even higher levels of state anxiety when subjected to our experimental conditions, thus resulting in lower GSR readings. While controlling for trait anxiety did not identify any effects on our outcomes, the effects of exaggerated state anxiety may have influenced our outcomes. In a study of the impact of emotion on cognitive load measurement using pupilometry and eye tracking (S. Chen & Epps, 2013), eye metrics including pupil dilation and blink rate were not affected by emotional arousal in task performance situations. The authors concluded that cognitive load was more influential than emotion in eye tracking and pupilometry measures of cognitive load. Whether this holds true for GSR measurement remains uncertain, but future studies measuring, and controlling for, state anxiety may offer some insight.

Future Directions

While VAS measurement of cognitive load may be superior to GSR in our experiments, the ultimate goal of this research is to identify a modality of cognitive load measurement that can provide reliable and accurate information to SBME educators in real time. This would provide SBME educators with data, enabling them to modify the simulated experience in real time, in response to the learner's cognitive load. Ultimately this may increase the educational efficiency of SBME. Theoretically, GSR measurement seems to have this potential, while VAS measurement does not. The VAS data were obtained during retrospective video review of their simulation performance, and thus, would not be available to educators in real time. It would be impractical to temporarily pause a simulated resuscitation exercise in order to have learners' reflect on their cognitive load. This would create an unnecessary interruption in the cognitive processes, and emotional engagement, used in SBME, and the added self-reflection may actually impose its own extraneous cognitive load on the learner. Furthermore, retrospective analysis of one's mental effort may be biased by success or failure. The outcome of the simulation exercise may unduly influence an individual's sense of self efficacy, and therefore their reporting of their mental effort.

GSR measurement is a non-invasive, minimally intrusive, and relatively inexpensive modality that can provide SBME educators with real time cognitive load data. Other useful, non-intrusive measures of cognitive load include behavioural cues that can be observed without disturbing learner activities. Specifically, as cognitive load increases, so do one's rate, energy, amplitude, and variability of speech. Linguistic and grammar changes can also be seen as cognitive load increases (F. Chen et al., 2012). These behavioural changes, in conjunction with other cognitive load measures such as GSR, may prove to be a valuable tool in the real time manipulation of cognitive load during SBME.

We suggest that further research, with appropriately-powered sample sizes, into GSR's role as a measure of cognitive load in dynamic environments, such as SBME, may validate its role as a reliable and minimally invasive tool to accurately measure the cognitive load of learners. This will ultimately allow educators to modify simulation scenarios in real time, using tenets of cognitive load theory, in order to maximize the educational value of each individual learner participating in SBME.

References

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review, 22*(4), 425-438. doi:10.1007/s10648-010-9130-y

Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability Generalization of Scores on the Spielberger State-Trait Anxiety Inventory. *Educational and Psychological Measurement, 62*(4), 603-618. doi:doi:10.1177/0013164402062004005

Barsuk, J. H., McGaghie, W. C., Cohen, E. R., O'Leary, K. J., & Wayne, D. B. (2009). Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit*. *Critical care medicine, 37*(10), 2697-2701.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276-292. doi:10.1037/0033-2909.91.2.276

Binstadt, E., Donner, S., Nelson, J., Flottemesch, T., & Hegarty, C. (2008). Simulator Training Improves Fiber-optic Intubation Proficiency among Emergency Medicine Residents. *Academic Emergency Medicine, 15*(11), 1211-1214.

Bohnen, N., Houx, P., Nicolson, N., & Jolles, J. (1990). Cortisol reactivity and cognitive performance in a continuous mental task paradigm. *Biological Psychology, 31*(2), 107-116.

Brydges, R., Hatala, R., Zendejas, B., Erwin, P. J., & Cook, D. A. (2015). Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Academic Medicine, 90*(2), 246-256.

Cant, R. P., & Cooper, S. J. (2010). Simulation-based learning in nurse education: systematic review. *Journal of Advanced Nursing, 66*(1), 3-15.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332. doi:10.1207/s1532690xci0804_2

Chen, F., Ruiz, N., Choi, E., Epps, J., Khawaja, M. A., Taib, R., . . . Wang, Y. (2012). Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(4), 22.

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology, 107*(3), 689-704. doi:10.1037/edu0000018

Chen, S., & Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer methods and programs in biomedicine, 110*(2), 111. doi:10.1016/j.cmpb.2012.10.021

Chen, S., Epps, J., & Chen, F. (2011, 2011). *A comparison of four methods for cognitive load measurement*.

Conway, D., Dick, I., Li, Z., Wang, Y., & Chen, F. (2013). *The effect of stress on cognitive load measurement.* Paper presented at the IFIP Conference on Human-Computer Interaction.

Ericsson, A., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review, 102*(2), 211-245. doi:10.1037/0033-295X.102.2.211

Fraser, K., Ayres, P., & Sweller, J. (2015). Cognitive Load Theory for the Design of Medical Simulations. *Simulation in Healthcare, 10*(5), 295-307. doi:10.1097/sih.0000000000000097

Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., & McLaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. *Medical Education, 46*(11), 1055-1062.

Gopher, D., & Braune, R. (1984). On the Psychophysics of Workload: Why Bother with Subjective Measures? *Human Factors, 26*(5), 519-532. doi:10.1177/001872088402600504

Hess, E. (1964). Pupil size in relation to mental activity during simple problem solving. *Science, 140*, 1190-1192.

Issenberg, B., McGaghie, W., Petrusa, E., Gordon, D. L., & Scalese, R. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical teacher, 27*(1), 10-28.

Issenberg, B., McGaghie, W. C., Gordon, D. L., Symes, S., Petrusa, E. R., Hart, I. R., & Harden, R. M. (2002). Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. *Teaching and learning in medicine, 14*(4), 223-228.

Jezova, D., Makatsori, A., Duncko, R., Moncek, F., & Jakubek, M. (2004). High trait anxiety in healthy subjects is associated with low neuroendocrine activity during psychosocial stress. *Progress in Neuro-*

Psychopharmacology and Biological Psychiatry, 28(8), 1331-1336. doi:https://doi.org/10.1016/j.pnpbp.2004.08.005

Kalyuga, S., Chandler, P.,&Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: recent developments and applications. *Psychoneuroendocrinology, 19*(4), 313-333.

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic medicine, 61*(2), 154-162.

Kohn, L., Corrigan, J., & Donaldson, M. (2000). *To Err Is Human: Building a Safer Health System*. Washington, DC: The National Academies Press.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394-421.

Lader, M. (1969). Comparison of Amphetamine Sulphate and Caffeine Citrate in Man. *Pshycopharmacologia, 14*(2), 83-94.

Landsberg, L., & Young, J. (2001). Physiology and Pharmacology of the Autonomic Nervous System. In E. Braunwald, A. Fauci, D. DKasper, S. Hauser, D. Longo, & J. L. Jameson (Eds.), *Harrison's Principles of Internal Medicine* (15th ed., Vol. 1, pp. 438-450). New York: McGraw-Hill.

Leblanc, V. R., Regehr, C., Tavares, W., Scott, A. K., Macdonald, R., & King, K. (2012). The impact of stress on paramedic performance during simulated critical events. *Prehosp Disaster Med, 27*(4), 369-374. doi:10.1017/s1049023x12001021

Marshall, S. P. (2007). Identifying cognitive state from eye metrics. *Aviation, space, and environmental medicine, 78*(Supplement 1), B165-B175.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 43-52. doi:10.1207/s15326985ep3801_6

McGaghie, W., Issenberg, B., Cohen, E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Academic medicine: journal of the Association of American Medical Colleges, 86*(6), 706.

McGaghie, W. C., Issenberg, S. B., Barsuk, J. H., & Wayne, D. B. (2014). A critical review of simulation‐based mastery learning with translational outcomes. *Medical Education, 48*(4), 375-385.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine, 6*(7), e1000097.

Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. (2012, 2012). *Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks*.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429.

Paas, F., & van Merriënboer, J. J. G. (1994b). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology,, 86*, 122-133.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Edcational Psychiologist, 38*(1), 1-4.

Paas, F., Tuovinen, J., Tabbers, H., & VanGerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71. doi:10.1207/s15326985ep3801_8

Paas, F., & Van Merrienboer, J. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 51-71.

Paas, F., & Van Merrienboer, J. (1994b). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology,, 86*, 122-133.

Paas, F., Van Merrienboer, J., & Adam, J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419-430.

Rodgers, D. L., Securro, S., Jr., & Pauley, R. D. (2009). The effect of high-fidelity simulation on educational outcomes in an advanced cardiovascular life support course. *Simulation in Healthcare: The Journal of The Society for Medical Simulation, 4*(4), 200-206.

Schaefer, J. J., 3rd, Vanderbilt, A. A., Cason, C. L., Bauman, E. B., Glavin, R. J., Lee, F. W., & Navedo, D. D. (2011). Literature review: instructional design and pedagogy science in healthcare simulation. *Simulation in Healthcare: The Journal of The Society for Medical Simulation, 6 Suppl*, S30-41.

Schreinicke, G., Hinz, A., Kratzsch, J., Hüber, B., & Voigt, G. (1990). Stress-related changes of saliva cortisol in VDU operators. *International archives of occupational and environmental health, 62*(4), 319-321.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). *Galvanic skin response (GSR) as an index of cognitive load*. Paper presented at the CHI '07 Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA.

Swanson, H. L. (1995). Effects of dynamic testing on the classification of learning disabilities: The predictive and discriminant validity of the Swanson-Cognitive Processing Test (S-CPT). *Journal of Psychoeducational Assessment, 13*(3), 204-229.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (Vol. 1). New York, NY: Springer.

Szulewski, A., Fernando, S. M., Baylis, J., & Howes, D. (2014). Increasing Pupil Size Is Associated with Increasing Cognitive Processing Demands: A Pilot Study Using a Mobile Eye-Tracking Device. *Open Journal of Emergency Medicine, 2014*. doi:10.4236/ojem.2014.21002

Szulewski, A., Gegenfurtner, A., Howes, D. W., Sivilotti, M. L. A., & van Merriënboer, J. J. G. (2016). Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. *Advances in Health Sciences Education*, 1-18. doi:10.1007/s10459-016-9725-2

Szulewski, A., Roth, N., & Howes, D. (2015). The Use of Task-Evoked Pupillary Response as an Objective Measure of Cognitive Load in Novices and Trained Physicians: A New Tool for the Assessment of Expertise. *Academic Medicine, 90*(7), 981-987. doi:10.1097/acm.0000000000000677

van Merrienboer, J., Kirschner, P., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist, 38*(1), 5-13. doi:10.1207/S15326985EP3801_2

Wayne, D. B., Butter, J., Siddall, V. J., Fudala, M. J., Linquist, L. A., Feinglass, J., . . . McGaghie, W. C. (2005). Simulation-based training of internal medicine residents in advanced cardiac life support protocols: a randomized trial. *Teaching and learning in medicine, 17*(3), 202-208.

Wayne, D. B., Butter, J., Siddall, V. J., Fudala, M. J., Wade, L. D., Feinglass, J., & McGaghie, W. C. (2006). Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. *Journal of general internal medicine, 21*(3), 251-256.

Zarjam, P., Epps, J., & Chen, F. (2011). Spectral EEG features for evaluating cognitive load. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2011*(Journal Article), 3841.

Zendejas, B., Brydges, R., Wang, A. T., & Cook, D. A. (2013). Patient outcomes in simulation-based medical education: a systematic review. *Journal of general internal medicine, 28*(8), 1078-1089.

Appendix A: Queen's University Health Sciences & Affiliated Teaching Hospitals

Research Ethics Board Clearance Letter

Queen's UNIVERSITY

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD (HSREB)

HSREB Initial Ethics Clearance

April 04, 2016

Dr. Eric Bruder
Department of Emergency Medicine
Queen's University

ROMEO/TRAQ: #6018059
Department Code: EDUC-013-16
Study Title: Manipulation of Cognitive Load in Simulation Based Medical Education
Co-Investigators: Dr. D. Klinger, Dr. D. Howes, Dr. J. Kirby
Review Type: Delegated
Date Ethics Clearance Issued: April 04, 2016
Ethics Clearance Expiry Date: April 04, 2017

Dear Dr. Bruder,

The Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board (HSREB) has reviewed the application and granted ethics clearance for the documents listed below. Ethics clearance is granted until the expiration date noted above.

- Questionnaire
- Recruitment Email
- Information/Consent Form

Documents Acknowledged:

- CORE Certificate – E. Bruder

Amendments: No deviations from, or changes to the protocol should be initiated without prior written clearance of an appropriate amendment from the HSREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

Renewals: Prior to the expiration of your ethics clearance you will be reminded to submit your renewal report through ROMEO. Any lapses in ethical clearance will be documented on the renewal form.

Completion/Termination: The HSREB must be notified of the completion or termination of this study through the completion of a renewal report in ROMEO.

Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information.

Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.

Yours sincerely,

Albert J. Clark.

Chair, Health Sciences Research Ethics Board

The HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations, Canadian General Standards Board, and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is qualified through the CTO REB Qualification Program and is registered with the U.S. Department of Health and Human Services (DHHS) Office for Human Research Protection (OHRP). Federalwide Assurance Number: FWA#:00004184, IRB#:00001173

HSREB members involved in the research project do not participate in the review, discussion or decision.

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD (HSREB)

HSREB Amendment Acknowledgment/Ethics Clearance

July 13, 2016

Dr. Eric Bruder
Department of Emergency Medicine
Queen's University

ROMEO/TRAQ: #6018059
Department Code: EDUC-013-16
Study Title: Manipulation of Cognitive Load in Simulation Based Medical Education
Review Type: Delegated
Date Ethics Clearance Issued: July 13, 2016

Dear Dr. Bruder,

The Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board (HSREB) has reviewed the amendment application and granted ethics clearance/acknowledgement for the following: documents listed below.

- Removal of the pupil dilation measuring device
- Revised email text
- Revised information/consent form

Yours sincerely,
Albert F. Clark.
Chair, Health Sciences Research Ethics Board

The HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations, Canadian General Standards Board, and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is qualified through the CTO REB Qualification Program and is registered with the U.S. Department of Health and Human Services (DHHS) Office for Human Research Protection (OHRP). Federalwide Assurance Number: FWA#:00004184, IRB#:00001173

HSREB members involved in the research project do not participate in the review, discussion or decision.

Appendix B: Sample Galvanic Skin Response Data

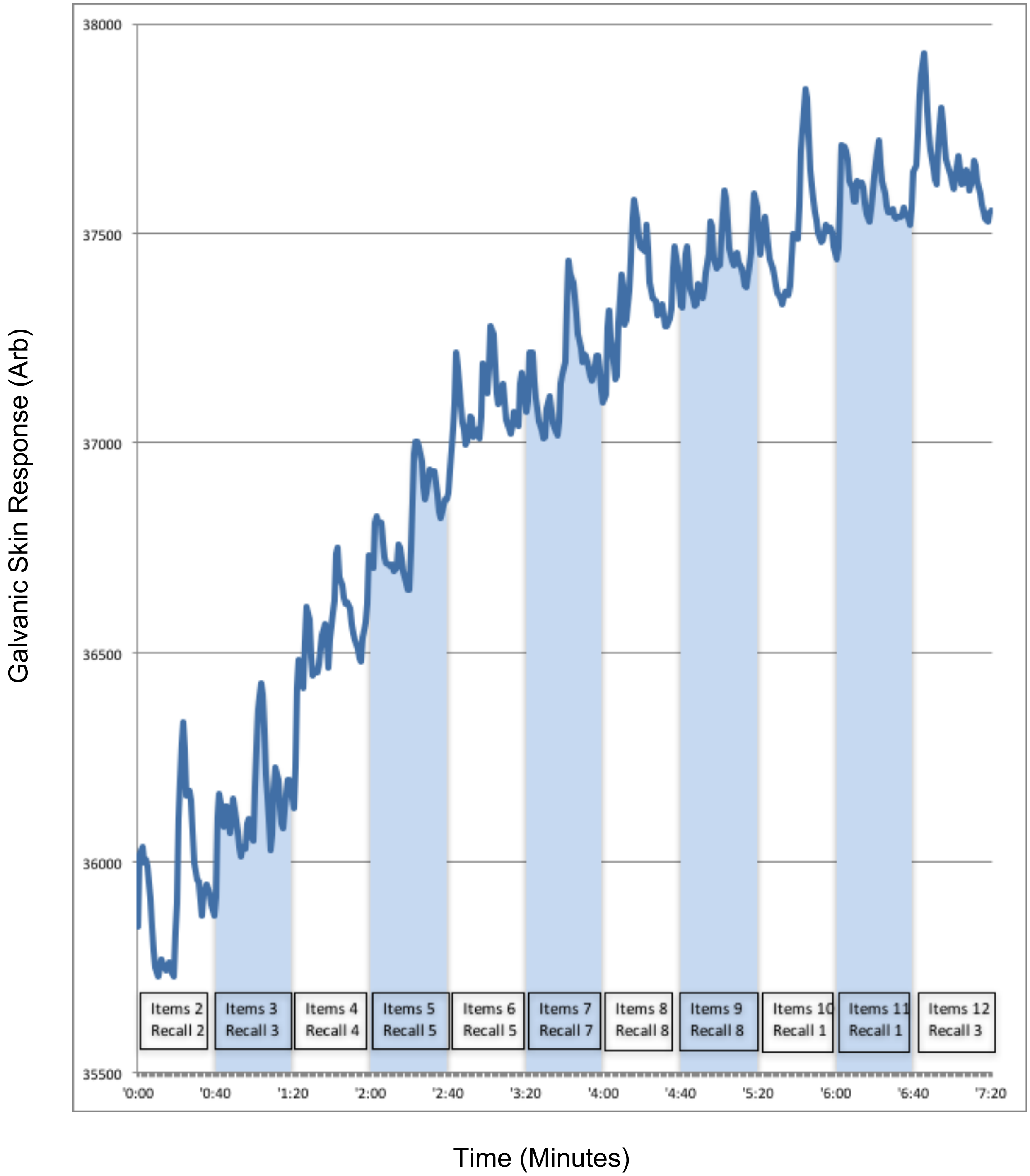

Figure B1. Sample galvanic skin response data during the visual spatial memory task. Each stage is identified by alternating shaded areas. Items=Number of items to recall in each stage. Recall = Number of correctly recalled items in each stage.

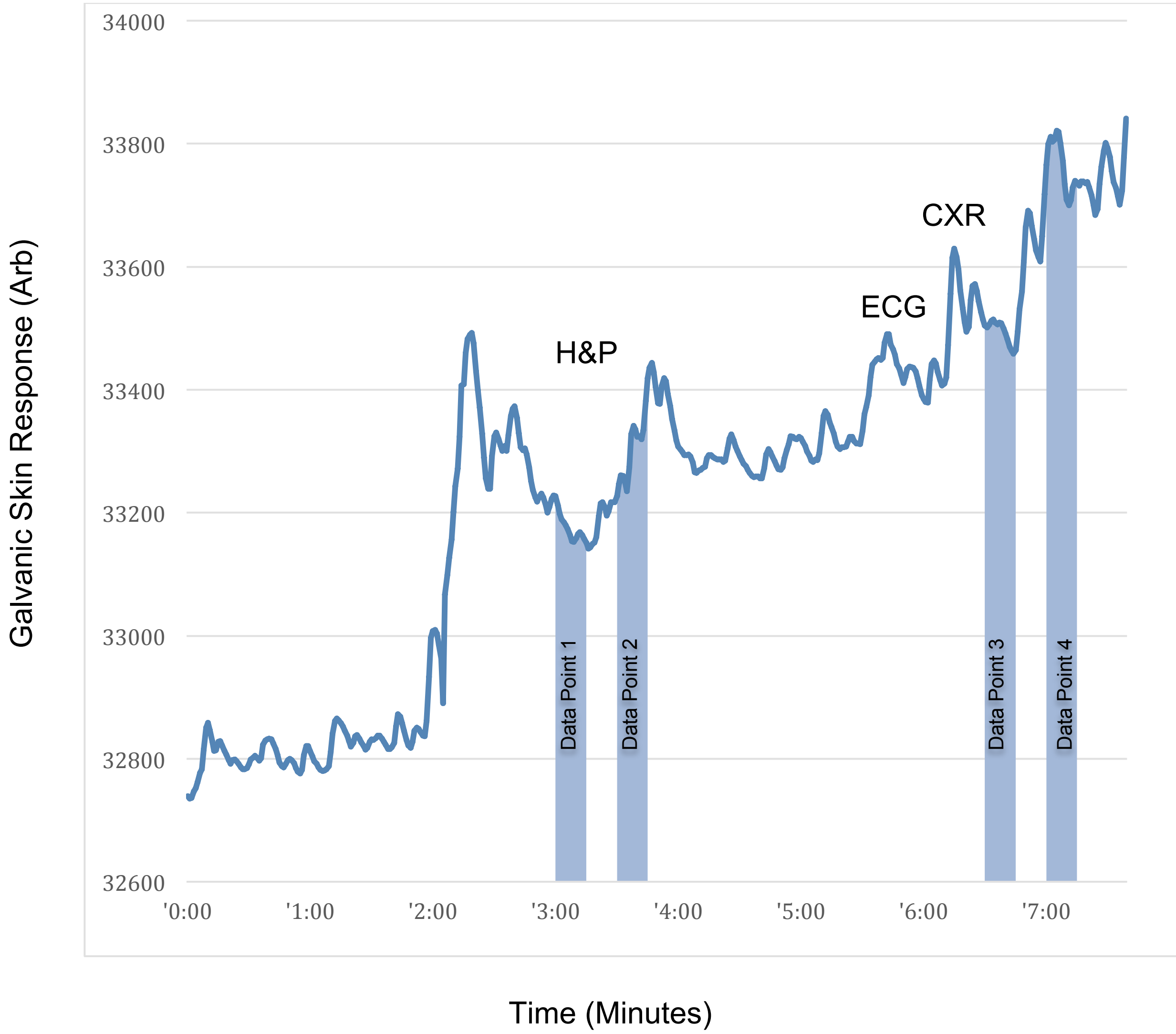

Figure B2. Sample galvanic skin response data during the first simulation exercise: Increasing element interactivity. Shaded areas depict data points that were analyzed. H&P=History and physical examination, ECG=Electrocardiogram interpretation, CXR=Chest X-Ray interpretation.

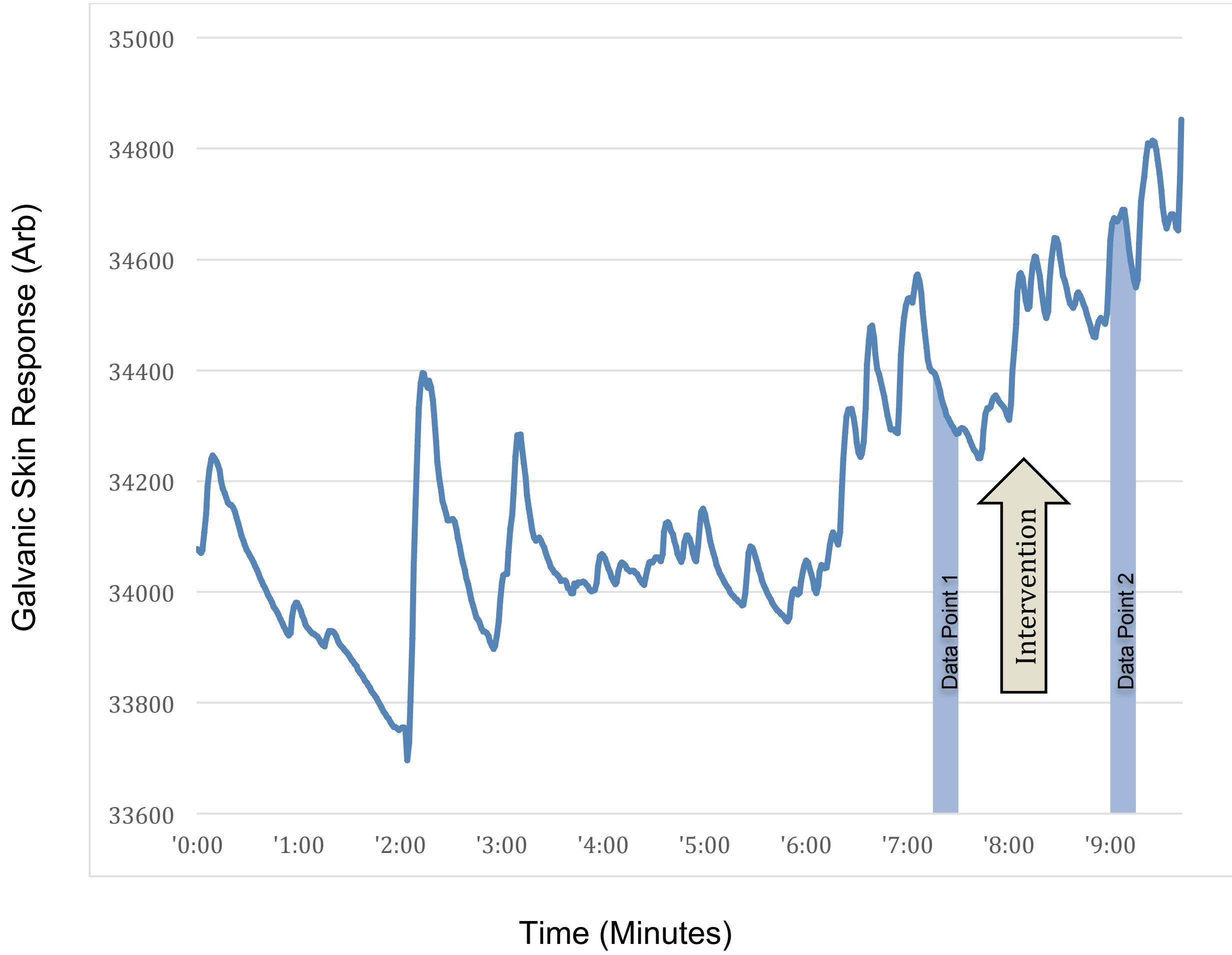

Figure B3. Sample galvanic skin response data during the second simulation exercise: Unpack schema. Shaded areas depict data points that were analyzed

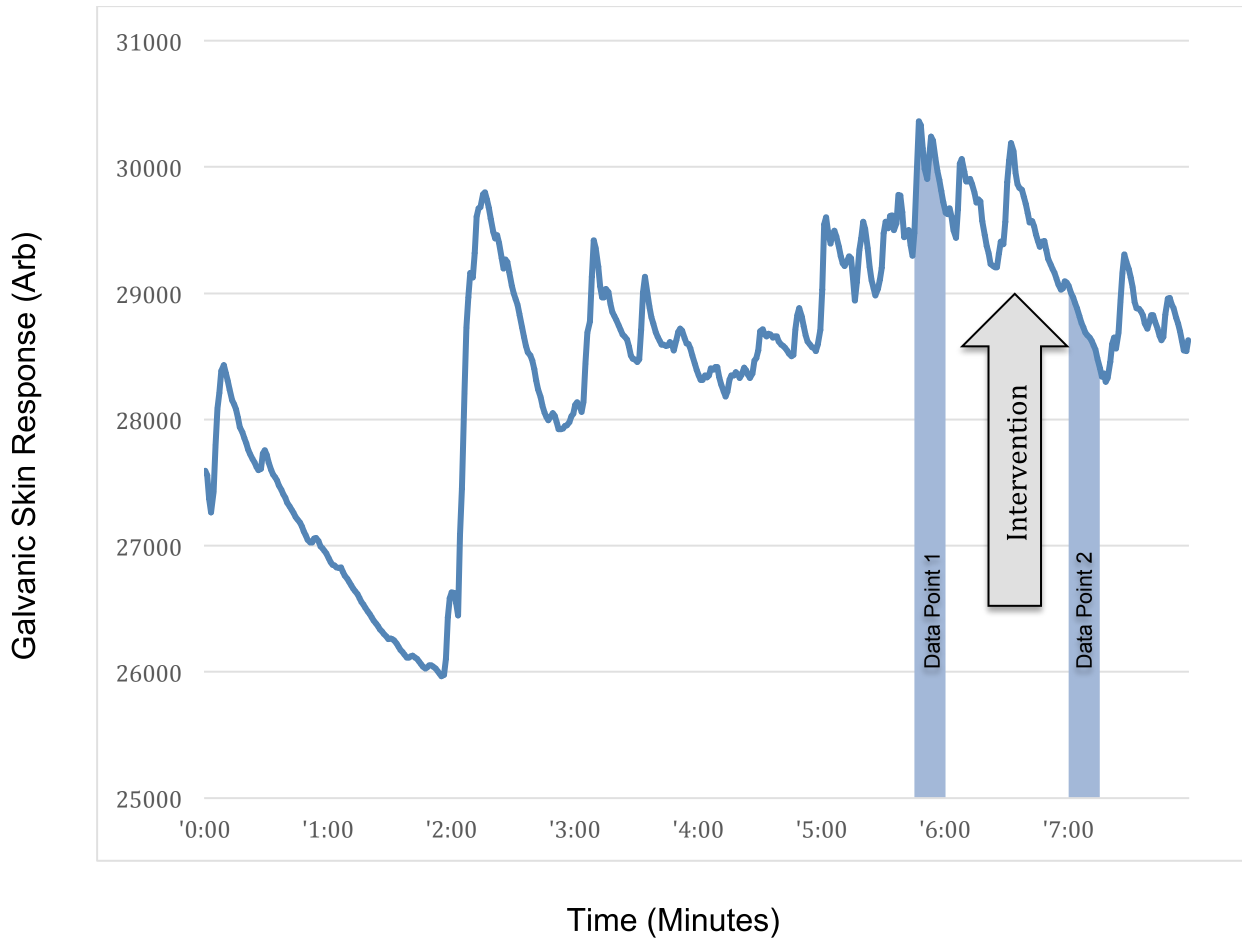

Figure B4. Sample galvanic skin response data during the third simulation exercise: Algorithmic approach. Shaded areas depict data points that were analyzed

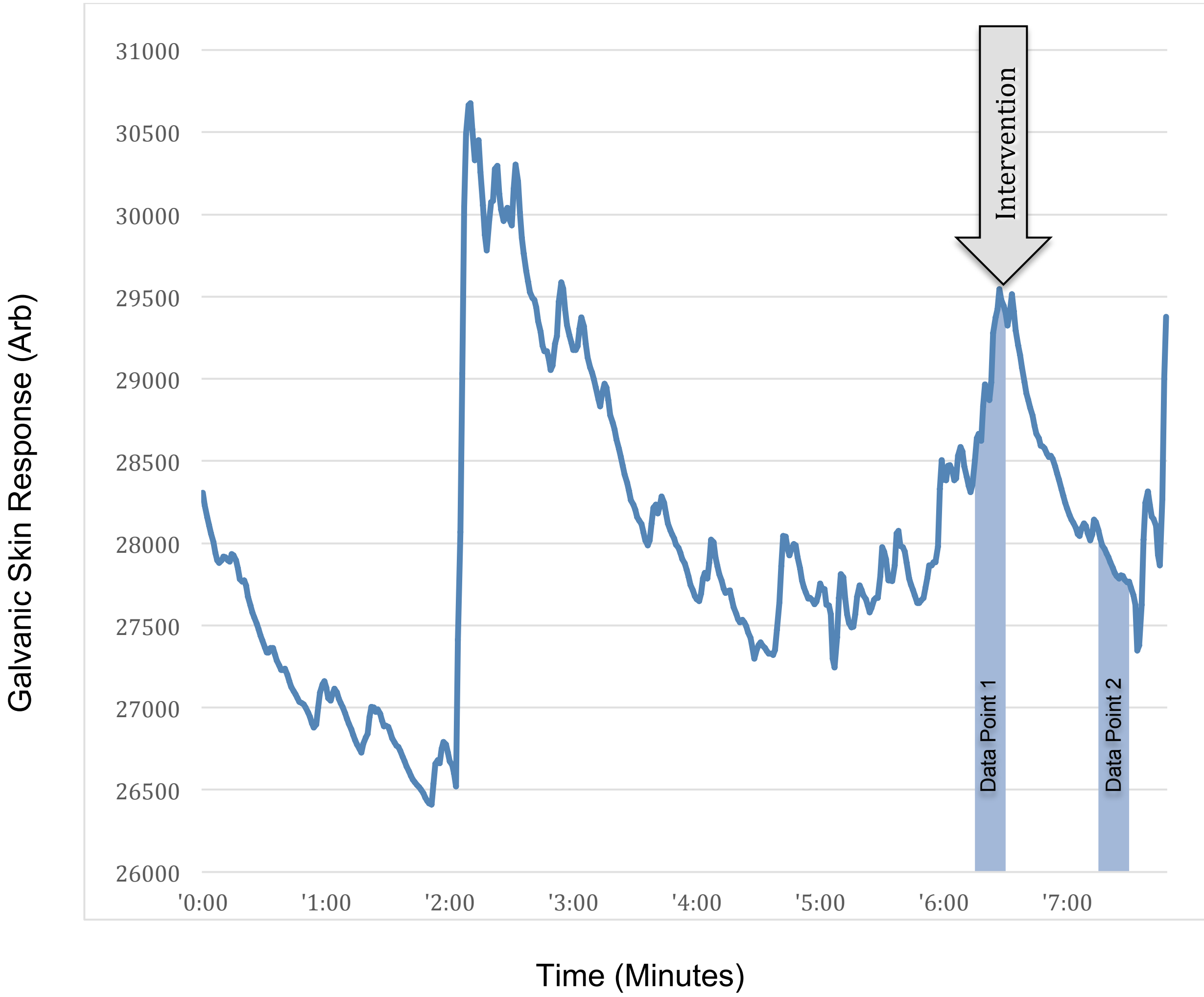

Figure B5. Sample galvanic skin response data during the fourth simulation exercise: Worked examples effect. Shaded areas depict data points that were analyzed