THE EFFECT OF USING SIMULATION WARS

ON EMERGENCY MEDICINE RESIDENTS' REASONING SKILLS

A Thesis Presented to the
Faculty of the College of Education
University of Houston

In Partial Fulfillment
of the Requirements for the Degree

Executive Doctorate in Professional Leadership
with an Emphasis in Health Science Education

By

Donna Mendez, MD

August 2016

THE EFFECT OF USING SIMULATION WARS

ON EMERGENCY MEDICINE RESIDENTS' REASONING SKILLS

A Thesis for the Degree

Executive Doctorate in ProfessionalLeadership
with an Emphasis in Health Science Education

by

Donna Mendez, MD

Approved by Thesis Committee

____
Dr. Sara McNeil, Chairperson

____
Dr. Bernard Robin, Committee Member

____
Dr. Guillermo Suarez, Committee Member

____
Dr. Margaret Watson, Committee Member

____
Dr. Robert H. McPherson, Dean
College of Education

August 2016

THE EFFECT OF USING SIMULATION WARS

ON EMERGENCY MEDICINE RESIDENTS' REASONING SKILLS

An Abstract of a Thesis
Presented to the Faculty of the College of Education
University of Houston

In Partial Fulfillment
of the Requirements for the Degree

Executive Doctorate in Professional Leadership
with an Emphasis in Health Science Education

By

Donna Mendez, MD

August 2016

Mendez, Donna. "The Effect of Using Simulation Wars on Emergency Medicine Residents' Reasoning Skills." Unpublished Executive Doctorate in Professional Leadership with an Emphasis in Health Science Education, University of Houston, Auigust 2016.

Abstract

In 2004, the Accreditation Council for Graduate Medical Education (ACGME) restricted the duty hours of residents training to become doctors in an effort to decrease medical errors. With less time for them to learn medicine at the patient's bedside, residents may not develop the clinical reasoning skills needed for diagnosing and treating patients. Simulation can remedy this, allowing residents to practice skills, to solve problems, and to develop clinical reasoning. Simulation is based on Kolb's Theory which emphasizes reflection of one's actions (Kolb, 1975), Lave and Wegners Theory based on socialization, visualization, and imitation (Hunmg 2002), and Social Constructivism Theory which is based on social interaction (Perkins, 2007). Simulation Wars, a clinical reasoning simulation, has rarely been incorporated into resident curriculum, but it has been shown to improve clinical reasoning skills (Young, Stokes, Denlinger & Dubose, 2007) (Yound, Dubose, Hedrick, Conaway, Nolley, 2007), (Hedrick & Young, 2008). In this study, the effect of Simulation Wars on In-service Board Scores and Global Rating Scale (GRS) Scores of Emergency Medicine (EM) Residents was analyzed. The results showed no significant difference in GRS scores except for Communication and Professionalism. There was no significant difference in the In-Service Board scores between the intervention group and control group. There was no improvement in the majority of subcategory topics on the In-Service board scores, but there was consistent improvement in some subcategories. In conclusion, Simulation

Wars has some applicability in providing Clinical Reasoning training in certain EM topics.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

Residents training to become doctors used to work many clinical hours in order to obtain the clinical experience needed to become a competent doctor, but now their working hours are restricted by regulations from the Accreditation Council for Graduate Medical Education (ACGME). In 2003, the ACGME implemented new rules that limited work hours for all residents (ACGME, n.d.). In contrast to the 110-hour weeks before the rules were created, current residents may work no more than 80 hours a week (Nasca, Day, & Amis, 2010). In addition, residents are restricted to no more than 24 hours of continuous work without taking a break, in contrast to the previous rule of 30 continuous hours. The aim of these standards was to promote high-quality learning and safe care in teaching institutions (Philibert, Friedman, & Williams, 2002) and to "protect patients from errors made by sleepy doctors" (Lloyd, 2013, p. 1 according to a study done at the University of Michigan Medical School (Sen et al., 2013). A landmark study (Landrigan et al., 2004) found that reducing medical residents' work hours during rotations in the intensive care unit resulted in a significant reduction in medical errors.

Working fewer hours can mean less learning time for residents. Restricting resident hours in the hospital can result in fewer opportunities for residents to perform procedures, manage medications, and/or gain experience with difficult clinical scenarios (Airee, Guirguis, & Mohammed, 2009; Eiser & Connaughton, 2011; Freed, Dunham, & Lamarand, 2009; O'Malley, Emanuele, Halasyamani, & Amin, 2008; Robinson, Mogensen, Grudinskas, Kholer, & Jacobs, 2005). With less time for residents to learn medicine at the patient's bedside, their knowledge base may suffer (Cohen-1

Gadol, Piepgras, Krishnamurthy, & Fessler, 2005; Immerman, Kubiak, & Zuckerman, 2007; Lieberman, Olenwine, Finley, & Nicholas, 2005; Mickelsen, Pankh & Persoft, 2011; Nuthalapaty, Carver, Nuthalapaty, & Ramsey, 2006). In addition, residents have reduced exposure to emergency resuscitations. In a study by Mickelsen and colleagues (2011), first year internal medicine residents were exposed to a 41% reduction in code blue events, between 2002-2008, as compared to previous years when there was not a reduction in hours. In a 2003 survey to internal medicine programs, residency directors felt that the restricted duty hours would negatively affect the quality of learning (Grześkowiak, 2006). Most of the program directors felt that residents ability to achieve competency in five of six required ACGME competency areas would be limited (Antiel 2010; 2011).

Simulation may contribute to the solution since learning could be enhanced with hands-on practice outside the hospital. Simulation dates back to the sixth centenary AD, with the game of chess used to simulate battlefield tactics (Reid, 2105). Simulators such as mannekins were first used in medical education in the 1960s for resuscitation and then anesthetic and clinical skills training (Bradley & Postlethwate, 2003). Simulation has become increasingly popular in medical education with use rising exponentially over the last 15 years.(Issenberg, McGaghie, Petrusa, & Scalese, 2005). Simulation in medical training has been shown to improve clinical skills leading to improved patient care practices and better outcomes. In a 2008 study, second year residents in internal medicine who were simulator trained were seven times more likely to adhere to Advanced Cardiopulmonary Life Support (ACLS) protocols than third-year residents who were ACLS recertified in the traditional manner (Wayne, Didwania, Feinglass, Barsuk, & McGaghie, 2008). In a study by Andreatta and colleagues in 2011 that used an instituted pediatric simulated mock code curriculum as a formal part of residency training, researchers noted an increase in cardiopulmonary arrest survival rates in the hospital over the 4-year period, with an increase from 33% to 50% within 1 year. Researchers have also reported better patient outcomes with simulation, such as reduction in catheter-related bloodstream infections (Barsuk, Cohen, Feinglass, McGaghie, & Wayne, 2009), postpartum outcomes such as neonatal hypoxic-ischemia encephalopathy (Draycott, et al., 2006), and brachial palsy injury (Draycott et al., 2008).

For the past decade, medical school curriculum has been under reform because of the expansion of biomedical sciences knowledge and pressure to add additional clinical experiences (Brauer & Ferguson, 2015). Traditionally, the first two years of medical school involve a lecture-driven format. The following two years include mainly clinical rotations with medical students learning through an apprenticeship model. Once a medical student graduates and starts residency they once again experience the apprenticeship model. (Bleakley, 2006). This apprenticeship model during residency involves, learning clinical skills at the bedside, including differential diagnosis, and treatment during residency. The apprenticeship model has also undergone a revolution (Ainley & Rainbird, 1978; Guile & Young, 2001) since previous apprenticeship models stressed ‘immersion’ learning by experiences gained simply through exposure. Previous apprenticeship models also employed the “sees one, does one, and teaches one’ experiences. An example of this would be a learner observing an attending performing a procedure such as suturing (sees one), then the learner would be able to perform it independently (does one) and then teach another learner (teaches one). This practice is no longer regarded as ethical or effective (Morgan & Cleave-Hogg, 2005; Ziv, Wolpe, Small, & Glick, 2003). Subjecting patients to potential harm as a consequence of training or lack of experience is not supported. New apprenticeship models, or "cognitive apprenticeships," developed in the early 1990s and stress that novices do not simply learn how to do the job as they gain expertise (Lave & Wenger, 1991). Novices should also learn how to "think" and "recount" the job (Bleakley, 2006, p. 154). With both the old and new apprenticeship models, direct contact of a novice with experts is required in order to learn important clinical medical concepts and patient care. This direct contact requires time at the bedside. However, the restricted hours a medical resident can work in the hospital has limited the amount of bedside teaching and learning. This has lessened patient care time which can prevent the resident from learning how to manage certain medical problems (Cohen-Gadol, et al., 2005; Immerman et al., 2007; Kupferman & Lian, 2005; Lieberman, Otenwine, Finley, & Nicholas, 2005; Mickelsen, et al., 2011). In addition, the reduction in clinical exposure has been implicated in the inability of junior doctors to recognize and manage critical ill patients (Mickelsen et al., 2011; Smith, Perkins, Bullock & Bion, 2007).

The medical school curriculum has undergone a change in hopes of making it more integrative. There has been a move away from extensive lecture time, a change from a teacher-centered approach to a learner-centered approach, and more integration of the basic sciences with clinical medicine in the first two years of medical school (Drake, 2014). An integrated curriculum moves through various systems in the body, such as cardiovascular, reproductive, gastrointestinal incorporating the basic sciences. The basic sciences includes such topics of anatomy, molecular mechanisms, physiology, biochemistry, and cellular mechanisms. The basic sciences material is linked to clinical problems often through patient cases. So the first year may consist of short courses providing fundamental information related to the basic sciences of the various systems of the body and then move to an integrated approach over the next three year where students see patients with these diseases. Since learning basic science is placed in the context of clinical and professional practice, the students have a reason to learn the information, making it more meaningful since they are applying the basic sciences to an actual patient (Pabst, Westermann, & Lippert, 1986).

During the last two years of medical school and through residency, students learn mainly through an apprenticeship model that involves clinical experiences with a patient. This has been more appealing to students than the lecture format of the basic sciences (Aljarallah, Hassan, 2015; Heckmann, Bleh, Dutsch, Lang, Neundorfer, 2003). Instead of learning from reading a text or attending a lecture, students learn from real patients at the bedside. This experience with live patient cases is more memorable compared to reading about medical problems in a textbook or attending a lecture on the subject (Heckmann et al., 2003). Several studies have compared interactive learning to traditional lecture ssand had improvement in knowledge acquisition as represented by test scores (Bulstrode, Gallagher, Pilling, Furniss & Proctor, 2003; Costa, van Rensburg, & Rushton, 2007; Papanna, 2013). Learning and retention also improved through the use of more interactive activities such as problem based learning (Louw, Eizenberg, & Carmichael, 2009; Sugand, Abrahams, & Khurana, 2010). Learning from a live patient does have its negative aspects, since it takes time for the learner and expert to do bedside teaching.

Research Question

Given the statement of the problem, the following research question was posed: Does teaching clinical reasoning skills using a simulation method of "Simulation Wars" affect Emergency Medicine Residents clinical reasoning skills as measured on In-Service Board exams, and Global Rating Scales (GRS) scores?

Context for the Study

This study occurred in a medical school in a major city located in the Southwest of the US. The data was In-service Board Scores and Global Rating Scale (GRS) Scores of Emergency Medicine residents in training.

Significant of the Problem

The study has profound clinical implications. Cognitive error contributes to 74% of diagnostic errors (Graber et al., 2005). Reducing diagnostic errors could greatly improve the practice of medicine and decrease unnecessary tests and invasive procedures ordered, as well as reduce unnecessary hospital admissions. With the inclusion of mechanisms such as simulation for the reduction of cognitive errors, there could also be a reduction in patient mortality and morbidity. It has been shown that the inclusion of mechanisms to reduce medical errors in anesthesia led to a 10-fold decrease in patient death rate (Ledue, 2010). Medical errors cost the U.S. economy almost $20 billion as reported by the Society of Acturies in 2010, with the average total cost per error of approximately $13,000. Reducing diagnostic errors could decrease the financial costs. With the ever rising cost of healthcare, reducing a medical errors which costs the US government millions of dollars would greatly benefit the US's deficit. Diagnostic errors also contribute to disability; reports show 10 million excess days are missed from work due to short-term disability due to errors (Ledue, 2010; Andel, Davidow, Hollander, Moreno, 2012). This same study found that $1.1 billion was the result of lost productivity due to related short-term disability claims. Reducing the disability rate that is caused by medical errors will enable those patients to work, enabling them to be financially independent and contribute to the national economy by their continuation of work.

Educational Value of the Study

The effect of Simulation Wars, as a way of teaching clinical reasoning skills, on In-Service Board exams and Global Rating Scale (GRS) scores has many applications. In-Service Boards and rating scales evaluate residents' clinical reasoning skills and showing that simulation wars can increase scores is evidence that simulation wars is an effective way of teaching residents. Other residency programs could incorporate Simulation Wars into their curriculum improving clinical reasoning skills and contribute to decreased diagnostic errors.

Limitations

The most prevalent limitation of the study was it's retrospective design. We were not able to randomly assign students to an intervention and control group so the GRS scores had different evaluators. The same group of EM supervising (attending) physicians were not in both groups because the control was from previous years than the intervention group. This was because some EM attendings left the job and new EM attending were hired. The groups may have been different in the way they evaluated EM residents accounting for no significant difference in GRS scores for those participating in the Simulation Wars. Another limitation was the Simulation War scores to evaluate how well the EM resident performed in the Simulation Wars were not consistently recorded. This was partly due to the design of the study, it being a retrospective study. Since this was the case, instead of comparing Simulation War scores with In-Service Board scores and GRS scores, we evaluated whether EM residents' participation in Simulation Wars would affect In-Service Board scores and GRS scores. Another limitation was the fact that this study occured at a single residency program at a single site affecting the generalizability of the results.

Definitions

Conceptual Definitions

Emergency Medicine (EM) Residents- EM residents are learners who have completed medical school and are now in postgraduate training.

Simulation Wars: Simulation Wars is a method of simulation where a group of residents are involved in different emergency patient scenarios and voice critical actions based on the scenario given. A critical action is considered an accepted standard of care for a clinical scenario. An example would be getting an EKG for a patient presenting with chest pain characteristic of a heart attack. Simulation Wars occurs in a controlled, reproducible environment outside the ER, such as in a conference room or lab. This can be done with individual learners or groups of learners. In this study, Simulation Wars was used by three to four residents, with one assigned as the leader and the others in the group as observers. Simulation Wars was started in 2011. Prior to this there was no type of clinical reasoning activity for EM residents. Simulation Wars occurred quarterly and included two to three scenarios each session. The scenarios include numerous body systems which match the testing material in the In-Service board exam.

In- Service Exam: An In-Service exam is a written multiple choice test given mid-year to all residents which includes topics on Emergency Medicine which are taught during lectures, assigned readings or at the patient's bedside. The questions are developed by the American Board of Emergency Medicine (ABEM). All residents take the same exam that is based on the knowledge and experience that a third year resident should have acquired. The test is taken online at a test-taking site outside the medical school. The Score is the percent of scored questions answered correctly. The number of scored questions vary from year to year. The Percent Rank indicates the percent of all residents in ACGME accredited EM residency programs, according to training level, who scored the same or lower than the resident of that year's exam.

Global Rating Scales Scores:

The GRS scores are used to assess resident progress and to produce the composite evaluations required by the Accreditation Council for Graduate Medical Education (ACGME). This is filled out by faculty who work with the resident monthly. There are the six medical competencies on which the resident is evaluated: 1. Patient care 2. Medical knowledge 3. Interpersonal communication and skills 4. Practice based performance 5. Professionalism 6. System based performance included for evaluation. The faculty scored the resident on a scale of 1-5, with 1 being the lowest and 5 being the highest was used from 2011 -2015. The scoring of a scale of 3; Above Expectations, Met Expectations and Below Expectations was used prior to 2011. The 5 Likert scale was transformed into the 3 level scale for purposes of analyzing the data.

Summary

This chapter discussed how residents' training time has been reduced to prevent medical errors. In the process, medical knowledge acquisition by residents has been felt to have suffered. Simulation can be the remedy for this problem, allowing residents to practice in the lab what they won't be exposed to at the bedside. Emergency Medicine is one of the specialties that requires clinical reasoning skills. Simulation Wars, a simulation method for learning clinical reasoning skills may prove to be helpful as measured by their In-Service board scores and GRS scores.

Chapter II

Review of the Literature

There are several learning theories that are incorporated into simulation. A discussion follows to demonstrate how different learning theories have a basis in simulation. This will help to connect our study to theories that have been incorporated in simulation, such as the Kolb's Experiential Learning Theory (ELT), Lave and Wegner Situated Learning Theory, and Social Constructivism Theory.

Kolb's ELT and Learning Cycle

Simulation-based training is closely aligned with the (ELT) and Kolb's Learning Cycle since a learner experiences the simulation by acting through the simulation and then reflects on one's actions. Through simulation, multiple experiences can be created to allow "reflection-in-action" (during the simulation) and reflection-on-action (critical review of one's own and other experience as a resource for learning) (Kolb, 1975). Kolb's Learning Cycle is composed of four main components that reflect the process by which learners are able to make meaning of an experience and use this newly gained information for decision making in the future. The cycle includes these stages: (a) concrete experience (b) observation and reflection, (c) forming abstract concepts and generalizations, and (d) testing implications of concepts in new situations (Miettinen, 2000). The cycle is continuous with the learner experiencing one step after another and, at times, engaging in them concurrently. The continuous flow through the phases differs greatly from the traditional form of medical education in that Kolb's Learning Cycle is neither educator focused nor does it concentrate on a lecture format. It allows the learner an opportunity to process an experience as a means of generating new ideas and concepts.

This cycle is clearly reflected in the methodology of simulation-based training described below.

Cycle step	Simulation-based training component
Concrete experience	Scenario—hands-on, sense of time pressure, multidisciplinary collaboration
Observation and reflection	Debriefing—facilitated by skilled instructors, open-ended questions posed to trainees
Abstract conceptualization	Debriefing—trainees work through questions posed by instructors in combination with video review to create generalizations
Active experimentation	Scenario—trainees are able to test newly formed concepts in the simulator in another hands-on scenario

Lave and Wegners Situated Learning Theory

Most medical errors occur due to systems failure rather than individual error while traditional medical education has focused on individual learning of delivery of care to individual patients, it has not focused on the importance of teamwork and the development of safe systems. Most medical errors are a consequence of system failures resulting from poor communication between professional and patient or among health professionals, or poor back-up systems, and so on (Institute of Medicine 2000). The team practice used in resuscitation simulation training, engages Lave and Wegners situated learning theory (Perkins 2007). Situated learning was first projected by Jean Lave and Etienne Wenger as a model of learning in a community of practice. This type of learning allows an individual (students/learner) to learn by socialization, visualization, and imitation. With situated learning, the learning begins with people trying to solve problems (Hung, 2002). People try to solve problems by exploring real life situations to find answers, or to solve the problems. This theory describes the participation that occurs when novice learners participate in a community of practitioners and gradually move towards full participation as mastery of knowledge and skills progresses (Bradley et. al., 2003). This is what occurs in simulation where a team is involved in a resuscitation moving from novice to expert.

Social Constructivism Theory

Learners who are performing a simulation such as a resuscitation scenario, one of the most common simulations in health care, work in small groups. There are assigned roles of group leader, medication nurse, and airway manager. The work group helps social interaction between learners, which is the basis of social constructivism theory (Perkins, 2007). This theory emphasize the importance of the social context of learning and the social interaction more than the individual interaction with the environment (Bradley et al., 2003). In this way the learner may extend their achievement within a “zone of proximal development” (ZPD) in collaboration with their peers and instructors (Atherton, 2005). The ZPD is defined as “the distance between the actual developmental level, as determined by independent problem solving, and the level of potential development, as determined through problem solving under adult guidance, or in collaboration with more capable peers (Vygotsky, 1978). The ZPD represents the amount of learning possible by a student given the proper instructional conditions (Puntambekar & Hubscher, 2005).

Simulation Based Education

Simulation-based education allows the learner to experience an event and participate in facilitated reflection using effective debriefing techniques, before experimenting again. The repeated performance of an intended cognitive or psychomotor skill in a focused domain, coupled with rigorous skill assessment, allows the learner to receive specific informative feedback and then to improve their performance in a controlled environment. Simulation-based education is best used to prepare learners for real patients. It allows them to practice and acquire skills needed to care for patients and speeds their progression along the learning curve from novice to expert (Perkins, 2007).

Simulation in Medical Education

There have been a number of studies on simulation and its application in medicine educating medical students, residents and fellows on technical skills (Cook, et al., 2011). Simulation was first implemented in the medical school curriculum for teaching technical skills such as suturing, placing chest tubes, intubating and placing central lines. Prior studies of simulation demonstrated an improved performance in the achievement of technical skills in multiple disciplines (Issenberg, McGaghe, Petrusa, Lee & Scalee, 2005; Isenberg et al, 1999). Simulation which has been shown to improve clinical skills has led to improved patient care practices and better outcomes. Simulation has resulted in an increased adherence to advanced cardiac life support (ACLS) guidelines when compared with traditional training (Wayne et al., 2008). Other examples of improved patient care practices include better management of difficult obstetrical deliveries (e.g., shoulder dystocia) (Draycott, et al., 2008), laparoscopic surgery (Seymour et.al. 2002) and bronchoscopy (Blum, Powers, & Sundarasan, 2004). Better patient outcomes with simulation have been reported in several studies using historical control groups that address reduction in catheter-related bloodstream infections, (Barsuk, et al., 2009) postpartum outcomes, (e.g., brachial palsy injury) (Draycott et al., 2008) and neonatal hypoxic-ischemia encephalopathy (Draycott et al., 2006) among newborn infants.

Simulation has advantages in simulating clinical situations; as teams can learn and practice the required interventions in a safe environment, and thus potentially improve patient outcomes when these situations actually occur in real situations. Simulation can also identify individual and team weaknesses (Daniel, Lipman, Harney, Arafeh & Druzin, 2008). Observations of many providers performing the same simulation can reveal the most common mistakes, allowing for development of appropriate and efficient curriculum for future training (Crofts, Fox, Ellis, Winter, Hinshaw & Draycott, 2008;Maslovitz, Barkai, Lessing, Ziv, & Many, 2007). Ultimately, simulation can help with the reduction of clinical experience due to mandated reductions in resident work hours (Johannsson, Ayida & Sadler, 2005;Macedonia, Gherman & Satin, 2003).

Simulation as Educational Modality in Medicine

Simulation as an educational modality has advantages over other educational options. In the domain of crisis management simulation can assess what trainees would actually do rather than what they write or say as in a multiple choice question exams or oral exams. This method demonstrates the “shows how” level of performance on the Miller’s pyramid (Savoldelli et al., 2006). This is where the learner shows how to apply the knowledge they have learned previously in demonstrating it in a simulation scenario. Simulation has also been shown to have higher learning outcomes than other types of learning for health professionals, including lectures, papers, small group activities, and videos (Cook et al., 2011). In a metanalysis of 92 studies on simulation versus other instructional methods by Cook and colleagues, they showed significant outcomes of knowledge, satisfaction, process skills (efficiency), and product skills (procedural success) when compared to other types of learning.

Simulation methods have been incorporated into Emergency Medicine residency curricula mainly for teaching technical skills. Simulation can assess the competency in a controlled environment, since the Emergency Room (ER) is a poor setting for learning in real -life situations due to: the uncertainty of the process and the patients' response; the complexity of the problem and possible confounding variables and simultaneous processes; time pressure; and stress. Very little didactic teaching takes place in the midst of a crisis, and in an emergency the learner is often moved to the observer role, as the instructor or more experienced clinician takes over. Emergency medicine is a field that relies on rapid clinical decision-making yet rarely do emergency medicine physicians receive formal training in clinical decision-making (Chapman, Car, & Aubin, 2005;Sandhu, Carpenter, Freeman, Nabors, & Oson, 2005).

Simulation and Medical Clinical Reasoning

The use of simulation for the development of clinical reasoning is in its infancy with few published articles (Bond et al., 2008; Hedrick et al., 2008; Young, DuBose, et al., 2007;Young, Stokes, et al., 2007). Simulation is a well-established approach to improving manual procedural skills, but has not yet been evaluated extensively in its ability to improve clinical reasoning skills. It also remains to be demonstrated that simulation can replace experience in actual practice. More recently, simulation has been incorporated into medical school and residency curriculum for clinical reasoning skills.

Both the Accreditation Council for Graduate Medical Education in the United States and the Royal College of Physician and, Surgeons of Canada (Carroere. Gagnon. Charlin. Downing, & Bordage, 2009) have requested residency programs to better assess and certify residents' key competencies in becoming qualified physicians, including clinical reasoning. Simulation can be a solution to decreasing diagnostic errors in the medical field. Simulation could take the place of the clinical experience where the residents can learn diagnosing and treatment on simulated patients in the simulation laboratory. This is helpful since there is limited time in the hospital for learning on live patients, with the mandated restricted residency hours. Thus, simulation is a way to help residents make fewer diagnostic errors and allows them to learn on simulated patients. Additionally, the use of simulation to study clinical reasoning has several advantages. Cases can be designed using a wide range of clinical situations instead of relying on the clinical material in a hospital. In addition, a sequence of cases of increasing complexity, and with different emphases, can be presented in a short period of time. The responses can be recorded in real time instead of retrospectively, and the reasoning being used at each decision point can be examined by asking the subject to describe his thought processes.

Simulation Wars as Educational Modality in Residency

Simulation Wars, is a simulation teaching method currently incorporated into few graduate medical education programs. A clinical scenario is presented by a supervising physician to a small group or individual learners. The vital signs, physical exam is provided on a patient presenting. The supervising physician then asks the residents as a group what laboratory tests they would order and they respond and receive a score for their answers. The patient may decompensate and need an intervention such as (cardiopulmonary resuscitation) CPR and the supervising physician will ask the group what action they would perform if patient is without a pulse and they also receive a score for their answer. The scores are summed and then another group of residents participate in the same scenario. The groups' score sums are compared with the one with the highest score winning a prize such as a gift card. This is a way of operationalizing clinical reasoning but unfortunately, there has been very limited study of its effectiveness. In the EM residency training, Simulation Wars, is a simulation of short clinical scenarios of patients in a controlled, reproducible environment where clinicians can assess and improve clinical decision making. At this point, only three articles exist describing the use of Simulation Wars in residency education. The first article, by Young et al., published in the American Journal of Surgery in 2007, studied general surgery and emergency medicine residents who participated in Simulation Wars. Each individual learner was placed in a room with a simulated patient and a nonmedical person who asked the questions. The learner was given a scenario that was written and was asked questions by a nonmedical person who had a script to ask questions such as "why do you want to do that?" The Primary Investigator (PI) for the project was present in an adjacent room but was unknown to the subject. The PI possessed a microphone and headphone that allowed him to communicate with the nonmedical person. Key tasks were determined by the investigators prior to investigation and validated by three experienced surgeons. The nonmedical person would ask the learner if he wanted an order made, and the learner was instructed to think out load as part of the evaluation of his reasoning skills. After the simulation, the session was transcribed and independently analyzed by the PI and 2 additional investigators. Reasoning was determined by using the methods of

Patel, Arocha, and Kaufman (1994). Forward reasoning was evidenced by hypothesis based statements (the patient has a fever, I believe they have pneumonia, therefore I will order a chest X-ray, (p.161). Backward reasoning was evidenced by data-based statement ("the patient has a fever, I will obtain a chest x-ray, leukocyte count, hematocrit, blood cultures and sputum cultures and examine these results and develop a diagnosis, 101". Errors were determined using the methods described by Cook and Woods (1994). Their study included 9 residents. Result of the study showed residents with more than ten weeks of Intensive Care Unit (ICU) experience made fewer cognitive errors in simulation wars than those with less than 10 weeks experience.

In a second article by Young, DuBose, Hedrick, Conaway and Nolley, published in Journal of Trauma, in 2007, they investigated whether Simulation Wars scores would be different according to training level. They hypothesized that the junior and senior medical students would be less able than interns and residents to detail the actions needed to assess, intervene and stabilize patients. Medical students and residents rotating on the trauma and surgical intensive care unit participated. Twelve scenarios were created to focus on basic floor emergencies and the scenarios were presented to groups of learners based on their educational level. Scores were assigned for clinical actions ordered. The scenarios were validated by two critical care and 3 emergency medicine attending physicians. The scenarios were presented to the attending physicians and their responses to the scenarios were graded and used as the expert group. Attending physicians were supervising physicians who have finished residency and their responses were considered to be expert. Scores for the medical students were assigned by two examiners, and the average of the grades in each area were used. The scores were a ratio of actual to possible correct responses in each section, and in the entire exercise. The initial evaluation consisted of actions such as going to see the patient, performing a physical examination, gathering data from the chart, and evaluation and correcting problems in airway breathing and circulation. The secondary evaluation consisted of ordering actions based on findings in the initial evaluation (laboratory evaluations, X-ray examinations, fluid boluses, etc.). The results of this study showed that on initial examination, 4th year medical students and upper level residents performed significantly worse than experts, and 3rd year medical students and 1st year residents performed similarly to experts. On the second evaluation all groups performed significantly worse than the expert group. This was unlike the prior study since this was done on groups of learners versus individuals, which has an advantage in that there is a group dynamic working as a team, which happens in real life during an emergency.

The third article on Simulation Wars was accomplished by Hedrick and Young in 2008. This study aimed at evaluating whether repeated simulation in the Simulation Wars format would improve performance. Participants included medical students and residents on the trauma and surgical Intensive Care Unit (SICU). Subjects were individually given a nursing report of an unstable patient and asked to verbalize how s/he would manage of the situation. Responses were transcribed and graded. They found medical students, as well as first and second year residents performed worse than experts. After participation in more than three war game sessions, trainee's scores were similar to experts. Subjects with the least amount of clinical experience demonstrated the most improvement.

The intent of the ACGME in restricting duty hours was not intended to limit clinical time; it was meant to improve safety in the hospital by ensuring fewer medical errors. In

2008, the Institute of Medicine (IOM) released a report on resident duty hours that discussed the benefit of reduced duty hours. This report noted that sleep loss, inexperience, workload intensity, inadequate supervision, poor handover practices and systemic factors contributed to the errors made by residents (Ulmer & Wolman, 2008). Per previous studies, medical errors were shown to occur in substantial numbers (James, 2013) and were more likely in those inexperienced clinicians such as residents versus supervising attending physicians. (Kohn, Corrigan, & Donaldson, 1999). Medical error has been defined as "the failure of a planned action to be completed as intended or the use of wrong plan to achieve an aim 'as cited in Pronovost, Thompson, Holzmueller, 2005, (p. 2). Lengthy clinical hours have been associated with fatigue and medical errors. Thirty-nine percent of internal medicine residents reported making at least one major medical error and attributed it to increased fatigue and sleepiness (West, Tan, Habermann, Sloan, & Shanafelt, 2009). Forty-one percent of resident-physicians attribute their most serious mistake due to exhaustion (White, et al., 2008). With restriced duty hours one would think that medical errors have decreased because the resident is more rested but this is not the case. Medical errors harming patients increased 15% to 20% among residents working shorter shifts compared with residents who worked longer shifts (Sen, et al., 2013). From the study by Sen and colleagues in 2013, the residents with shortened duty hours reported the same amount of depressive symptoms with these depressive symptoms linked to more medical errors and poorer clinical performance (Farheinkopt et al., 2008) (West et al., 2006). In Sen's study, they found that interns with depressive symptoms reported medical errors 5.3% of the time, at almost twice the rate as non-depressed interns (17.8%). In 2003, a study of internal medicine residents who were under the restricted duty hours had no significant change in the number of medical errors when compared to those not on the restricted duty hours, as measured by inpatient mortality and patient length-of-stay in the hospital (Anupam, Schoemaker, & Bhattacharya, 2014). So even with the restricted hours, medical errors have not decreased. Since there has not been a reduction in medical errors with restricted hours we need to incorporate or improve the residency curriculum so medical errors decrease. This could happen with the incorporation of critical reasoning simulation exercises.

Medical Errors

During the last decade, there has been a tremendous upsurge of attention to medical errors (Aspden, Wolcott, Bootman, & Cronenwett, 2006). In a recent article, medical error was felt to be the third leading cause of death in the United States (Makary, 2016). The Emergency Department (ED) has been characterized as a ‘‘high hazard health care setting” (Leape, et al., 1991). The Harvard Medical Practice Study (1991) reported that 3% of adverse events uncovered occurred in the ED. Medical errors in the ED can be caused by a number of factors, including 1) the environment itself, 2) variety of patient problems and treatment options 3) time critical decisions, and 4) rapidly shifting priorities (Brennan, 1991).

The ED Environment. The ED environment is characterized by multiple interruptions in the workflow of the health care workers, interrupting their train of thought resulting in flawed or delayed tasks. In fact, it has been reported that an interruption occurs 6.9 times per hour in a hospital ED (Fairbanks, Bisantz, & Sunm, 2007). The ED physician whose attention is constantly shifting from one item to another may not be able to formulate a complete and coherent picture of the presenting patient.

EDs have environments that are constantly changing since the patient rarely stays in one location: the patient may be in an ED room, then moved to a hallway, or radiology area or to a procedure room. These factors - multiple interruptions and constantly changing environment - allow errors to occur.

Variety of Patient Problems and Treatment Options. The workload in the ED is uncontrolled and unpredictable with intermittent time-critical tasks such as performing cardiopulmonary resuscitation (CPR) or an intubation of a patient. The most high-risk patients pass through an ED on their way into the hospital, requiring more individual procedures and decisions, and are therefore exposed to more possibilities for error. Inconsistend Arrival of Patients. Another factor contributing to medical errors in the ED is the fact that patient visits are unscheduled with irregular peaks and troughs in the number of patients and acuity levels of illness and injury. The ED staff may be less attentive during slow peirosds or moe stressed during busy periods contributing to medical errors.

Rapidly Shifting Priorities. Emergency physicians must constantly shift their priorities to optimize care and must decide which task should be performed first since multiple tasks are encountered simultaneously. In addition to the multitasking that occurs, there is rarely a downtime to clean up or to restore order completely in the ED. Diagnostic errors are a widespread problem, the magnitude of which is largely unknown because researchers do not have a valid mechanism to measure such errors. According to a report by the Institute of Medicine in 2015, it is estimated that 5 percent of U.S. adults who seek outpatient care each year experience a diagnostic error. Postmortem examination research spanning decades has shown that diagnostic errors contribute to approximately 10 percent of patient deaths, and medical record reviews suggest that they account for 6 to 17 percent of adverse events in hospitals. This surpasses the annual number of deaths due to heart attacks and motor vehicle accidents (Ainley, et al, 2013). Cognitive factors such as thought-process errors or thinking mistakes contribute to up to 74% of diagnostic errors in internal medicine (Graber, Franklin, & Gordon, 2005) and can lead to incorrect diagnoses, treatments, or both (Stiegler, Neelankavi, Canales, & Dhillon, 2011). These diagnostic errors may be caused by inadequate knowledge, faulty data gathering, inaccurate clinical reasoning and erroneous verification of diagnostic hypothesis (Graber, Gordon & Franklin, 2002). Cognitive errors are more prevalent than medication errors or knowledge deficiencies. Knowledge deficiencies are insufficient knowledge of the signs and symptoms of a disease (Graber, 2005: Shojania et al., 2003). The tendency for cognitive errors to occur frequently is based on innate and learned cognitive processes that are usually unconscious and unrecognized as risks to patients. There is general consensus that cognitive errors occur because clinicians rely too much on pattern recognition and are not sufficiently thorough and reflective so they can search for additional diagnoses (Norman, 2009). Perhaps the most important approach for eliminating cognitive error is to continually force oneself to examine his/her thought processes (Croskerry, 2005).

Critical Thinking Training. One of the proposed strategies to reduce cognitive error is critical thinking training in medical school. Critical thinking is the most important key in any educational program and especially in the medical profession where diagnosing a patient correctly is of utmost importance because of the negative consequences or even death if not diagnosed correctly (Papp et al, 2014). The American

Philosophical Association (APA) (1990) defined critical thinking as “purposeful, self-regulatory judgment that uses cognitive tools such as interpretation, analysis, evaluation, inference, and explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations on which judgment is based” (p.10). The elements of critical thinking include what has variously been described as clinical judgment (Feinstein, 1967), medical decision making (Sos, Blatt, Higgins & Morton,1988), clinical decision making (Eddy, 1996), and clinical reasoning (Weinstein et al., 1980). After decades of relative neglect, critical thinking became more popular in education in the 1980s leading to its incorporation into elementary, K-12 and college courses. Since there are various terms for critical thinking we will refer to critical thinking as clinical reasoning in the rest of this paper. In medical education, clinical reasoning is viewed as the process whereby a physician will think of the possible diagnoses for a patient’s problem and will think through them and refer to medical evidence to help with the process. Through the process, errors in clinical decision-making can be reduced.

Clinical reasoning is essential for competency in medicine. Clinical reasoning is based on the The Dual-Process theory (Croskerry, 2009). It proposes that physicians use two types of processes when thinking through a case. The first process is ‘‘intuitive,’’ and refers to the non-analytical processes. It is triggered and proceeds automatically, with no conscious effort (Charlin, Tardif , & Boshuizen 2000;Pelaccia, Tardif, Triby, & Charlin 2011). It allows physicians to rapidly generate diagnostic hypotheses using whatever information they have at hand and is done mostly from pattern recognition (Marcum, 2012;Norman, Young, Brooks, 2007). The other process of clinical reasoning is characterized as analytical reasoning. Analytic reasoning uses deductive reasoning strategies where information is gathered and applied. (Croskerry, 2009;Pelaccia, Tardif, Triby, & Charlin, 2011). With increasing clinical experience, learners do not simply gain new knowledge in an additive manner but they reorganize existing knowledge in a way that allows them to become more efficient and effective in diagnosing and treating new patents who have features similar to those of previous patients (Bordage G, 1994;Bowen 2006;Patel, Groen, & Patel, 1997;Schmidt & Boshuizen 1992). This is in line with Kolb's theory which suggests in his experiential learning theory, that knowledge is strengthened through reflective observation, active experimentation (e.g,, in a simulation environment), and abstract conceptualization (Bereiter & Scardemalia 1993;Kolb, 1983).

The stages of clinical reasoning go from novice to the advance learner. A novice learner is less developed and attempts to gather essential and accurate information about the patient's history and physical examination. Novice learners may not successfully recognize the clinical presentation and/or may give less important features that are not as important more attention, leading to suboptimal clinical reasoning which leads to the wrong diagnosis. The novice learner tends not to change their early diagnosis even as more clinical information is gathered (Schmidt, Norman & Boshuizen, 1992; Schmidt & Rikers, 2007;).

The novice learner goes to an intermediate level with even further clinical experience, with the learner organizing knowledge as scripts in which the characteristic features of specific illnesses form clinical patterns in memory (Schmidt et al. 2007)(Schmidt et al., 1990). A script is an organized network of knowledge. These scripts are formed from knowledge and experience. At this state, the intermediate learner arrives at a diagnosis through matching the patient’s clinical picture to his illness script for this disease, which may include physical exam finding and laboratory findings. These scripts can be used to compare and contrast diagnostic possibilities by comparing clinical patterns of disease presentations. The learner continues to refine and remodel these scripts as they encounter new patients. Over time, they become robust presentations of diseases, each one with discriminating features of illnesses that even become context dependent. (Schmidt, et al.,2007). Memories of individual patients seen in the past add specific situational information to these scripts, transforming scripts into "instance scripts" unique to individual patients (Schmidt, 1992). The intermediate learner tends to change their diagnosis as more information such as labs and history become available but their therapies tend to treat the features of a diagnosis rather than a unified diagnosis.

The advanced learner, through his/her clinical experience, attains the knowledge structures formed previously in the development of clinical reasoning. The advanced learner is not just using pattern recognition from previous cases they have seen. They are also engaging in analytic reasoning and using their understanding of the underlying causal mechanisms of disease and their pathophysiologic consequences to compare and contrast the discriminating features of the diagnoses they are thinking about (Patel, 1997; Schmidt, 1992). Advanced learners tend to develop quite advanced and narrowed diagnostic hypotheses early in a case and use of subsequent history, tests and physical exam findings to confirm their initial diagnosis. Their therapies tend to be focused and specific, based on a unified diagnosis (Schumaker, 2014). Since learners go through sequential stages of clinical reasoning we chose to look at not just one level of residency training but first, second and third level residents and how Simulation Wars affected these skills as one progressed during residency.

Emergency Medicine (EM) is one of the specialties that most frequently utilizes clinical reasoning. The ability to make rapid clinical decision is essential in the practice of emergency medicine. (Chapman, Car, & Aubin, 2005). Experienced emergency physicians often make complex clinical decisions with little medical information and limited time. They seldom summarize and evaluate their clinical decision making pathway (Xu, Xu, Yu, Ma, & Wang, 2012). There is little literature involving clinical reasoning in the chaotic ED. Likewise, there is limited evidence that simulation can improve clinical reasoning skills. There are only a total of three studies showing simulation can improve clinical reasoning skills in EM residency training (Hedrick & Young, 2008;Young, DuBose, Hedrick, Conaway, & Nolley, 2007;Young, Stokes, Denlinger, & Dubose, 2007). These studies were performed in the ICU and Emergency Department and had positive results with improvement in diagnostic skills. The Emergency Department is a logical place to teach clinical reasoning skills since the Emergency Department is a fast moving, multitask environment where diagnosis are made and also mistakes can occur. Mastering clinical reasoning skills can improve the accuracy of diagnoses which is imperative in the ED where the majority of diagnoses are made.

There are several methods that can be utilized to measure clinical reasoning skills. Medial resdients are typically evaluated with classic written examinations, GRS scores, patient satisfaction surveys and objective structured clinical examinations and simulation. The GRS scores are the most frequently used by Graduate Medical Education (GME) residency programs. The GRS scores evaluate the resident on variables that include medical knowledge and behaviors by the faculty that observed them while working the

ED. The six ACGME competencies for which residents are graded per the GRS scores include; medical knowledge, patient care, system based practices, communication and professionalism (Littlefield , Paukert , & Schoolfield, 2001;Gray, 1996). GRS scores by faculty have been viewed as a better measure of the residents' ability to apply knowledge than other methods (Haurani, 2007). Cognitive knowledge is measured by test scores such as the In- Service board scores and global rating scores but mainly with the In-Service Board scores. A study by Krishnamurthy in 2009, showed In-Service scores did not relate to decision making parameters. Our study included both the GRS scores and In-Service Board scores in order to include a measure of both medical knowledge and the application of this knowledge.

The In-Service board scores and GRS scores attempt to measure competency to ensure residents are more competent as they move from their first year to last year of residency. Competency in medicine is defined as well learned skills that, once acquired, may be used repeatedly on many patients. This should also involve the capacity to respond appropriately to multiple, simultaneous symptoms and medical events that can further complicate a patient's morbidity (Meterissain , Zabolotny, Gagnon, & Charlin, 2007).

In service EM board scores have been shown to be positively correlated to final certification board scores which evaluate clinical competency (Hamlin, Cutle, Hafner,&Thompson 2008). Final certification board scores correlate with the Continuous Certification (Concert) which is taken every 10 years by practicing emergency medicine physicians (Munger, 1982;Hamlin, 2008). The Concert examination that measures clinical knowledge and how it is applied in medical practice and must be passed every 10 years to maintain certification. The ConCert examination measures diagnostic reasoning by using psychometrically validated items. (Holmboe, Lipner, & Greiner, 2000). We did not have final certification board scores available to us for this study. Since In-Service Board scores were available and since it has been shown they correlate with clinical competency as measured by the final board scores and Concert, In-Service Board scores served as a measure of clinical competency.

Statement of the Problem

Emergency Medicine (EM) is one of the specialties that most frequently utilizes clinical reasoning since the ability to make rapid clinical decisions is essential in the practice of emergency medicine (Chapman, Car, & Aubin, 2005). Experienced emergency physicians often make complex clinical decisions with little medical information and limited time; they seldom summarize and evaluate their clinical decision making pathway (Xu, et al., 2012). There is little literature involving clinical reasoning in the chaotic ED. Likewise, there is limited evidence that simulation can improve clinical

Chapter III

Methodology

The Research Question

Does teaching clinical reasoning skills using a simulation method of "Simulation Wars" affect Emergency Medicine Residents clinical reasoning skills as measured on In-Service Board exams and GRS scores? The question proposed will investigate whether "Simulation Wars," a unique method that has not been employed infrequently to teach clinical reasoning skills, influences resident In-service Board Scores and GRS Scores. Clinical reasoning skills in the Emergency Room are critical since this is where physicians need to make a quick diagnosis so appropriate treatment can be administered. This prompt diagnosis and treatment is vital to the patient's survival.

Variables

The independent variable in this project was whether students participated in the Simulation Wars, which was be used to teach clinical reasoning skills. Since some of the residents received the training, we compared the scores of those undergoing the training to those who were untrained. The dependent variable was the residents' scores on the In-Service Board exams and GRS scores. Standardized assessments to access clinical reasoning skills included the In-Service exam and GRS scores.

Moderators

Moderators are those variables that affect or modify the basic relationship between the primary independent variable and dependent variable. There are several moderators that may influence emergency medicine resident scores on In-Service Exams and (GRS) scores. The level of training will be a possible moderator, with third-year residents potentially doing better than first or second year residents. Also the number of Simulation Wars sessions the residents attended could possibly influence In-Service Exams and GRS scores. There are other variables such as the residents' personal experience with a similar patient in the ER or in a previous job which could probably improve their scores. An example of these situations would be a resident having treated a patient previously with an aortic aneurysm and when given this same scenario on In-Service Boards, his score will be higher due to learning of the clinical presentation and treatment options based on this prior experience rather than the knowledge acquired during the simulation session.

Mediators

Mediators are variables that attempt to explain the relationship between two other variables. Suspected mediators will be whether the resident is a good standardized test taker, his/her anxiety level, and his/her confidence level.

Validity and Reliability

The scenarios and questions on the scenarios used for Simulation Wars come from the Council of Emergency Medicine Residency Directors (CORD). The scenarios and questions from CORD are from 44 Emergency Medicine residency programs directors; each scenario has gone through a peer review process. These questions have been validated by an expert group of physicians ("CORD," n.d.). The In-Service Exams, Oral Boards, and Certification Boards are reliable and valid, as they were formulated by the American Board of Emergency Medicine (ABEM) (Munger, Krome, Maatsch, & Podgomy, 1982Bianchi, Gallagher, Korte & Ham,2003;). The Global Rating Scales used for EM residents' evaluations for milestones are valid and reliable having undergone psychometric testing (Bartlett, et al., 2015).

Research Design

The Quasi-Experimental design was used since this was a retrospective study without a random assignment given while administering the Simulation Wars. The main comparison was with controls, who were residents who didn't participate in the Simulation Wars, and the intervention group, who were involved in the Simulation Wars. The preexisting data was separated out and analyzed into separate subgroups, with first year residents compared to the second year and third year residents. The effect of their involvement in Simulation Wars and their In-Service Board Exams and Global Rating Scales Scores were also measured.

Participants: The participants were Emergency Medicine (EM) residents training at The University of Texas Health Science Center at Houston. There was a historical control group of EM residents from years 2009-2011, whose In-Service Board Exams and Global Rating Scales Scores were compared to those EM residents involved in Simulation Wars (cases/intervention) during the years of 2012-2015.

Characteristics of the Participants: Participants consisted of first year, second year and third year EM residents. The whole group of residents was made up of approximately 60 residents and there were four (simulation) wars per year. The residency program usually enrolls more males than females in each group (those involved in Simulation Wars and those not involved in Simulation Wars). There were some residents in both groups who were nurses or paramedics prior to going to medical school so they may have had more clinical knowledge compared with those who were not in the medical profession in the past. All of the residents were US citizens and had passed the United States Medical Licensing Exams One and Two within two tries. Residents attended either a US or nonUS medical Schools or Osteopathic medical school.).

<u>Intervention (Treatment):</u>

Simulation Wars

Step 1. Students were in a large classroom for Simulation Wars. The attending physician (teacher) presented clinical scenarios to four residents who worked as a group with an assigned leader similar to a resuscitation code scenario. The scenarios are relevant ED patient presentations, such as a Heart Attack, Stroke, or Aortic Aneurysm. Scenarios in Signs and Symptoms, Abdominal/Gastrointestinal, Cardiology, Environmental, Head Eyes Ears Nose and Throat, Systemic Infections, Musculoskeletal, Obstetrics/ Gynecology, Psychology, Thoracic, Toxicology, Trauma and Nervous System were given during the study period. The attending physician presented the four residents the patient's vital signs and physical exam results. All the available supplies for intravenous line, intubation, and intravenous fluids were available to the group to simulate their use on the patient

Step 2. The faculty member asked the four residents "what you would do next?" Examples of this were for a patient presenting with chest pain that is typical of heart attack. The first step would be to place oxygen, administer nitroglycerine, morphine and perform electrocardiogram (EKG). The placement of the patient on oxygen, administration of nitroglycerine, morphine and performing an EKG are critical actions. The patient may decompensate and need pacing and the residents should recognize this with the vital signs and the EKG given.

Step 3. Another group of four residents who were unaware of the scenario given to the first group came into the classroom and was presented the same scenario.

Step 4. After the scenarios were finished, the attending physician then gave a short summary of the scenario including important points of the patient's diagnosis and treatment. The attending also went over the correct critical actions they should have ordered.

Data Collection Procedure

The project was given IRB approval from The University of Texas Health Science Center-Houston Medical School and The University of Houston prior to data being analyzed. Once the two IRBs were approved a data base was set up to collect EM resident demographics, In-Service Board exams and GRS scores. Coding by number was done to enter data so a first year resident was coded one, a second year coded as two and third year coded as three. The race, age and gender were also collected and coded. Other data such as the resident being employed as a healthcare worker in the past was recorded and coded. Board and Global Rating Scores were entered as individual scores.

Data Analysis Procedures

For comparisons between the historical control group and the intervention group (cases) of continuous variables, such as the In-Service Board exams and GRS scores, were evaluated using the Independent t test. Continuous variables were analyzed with the t-test if normally distributed or a parametric test such as the Mann-Whitney if non-normal distribution was found. Nominal variables were analyzed by using the Chi^2 test or Fisher's exact test.

Summary

The problem for residency training is limited time to learn at the bedside due to the restricted hours enforced by the ACGME. Simulation Wars has been shown to improve clinical reasoning skills in residents. We analyzed demographic and scoring data to investigate whether learning clinical reasoning skills with Simulation Wars helped improve In-Service Exams and GRS Scores.

Chapter IV

Results

Introduction

The purpose of the study was to determine the outcome of incorporating Simulation Wars as an instructional tool in clinical reasoning to Emergency Medicine Residents and the effect of this instructional tool on In-Service Board exams and Global Rating Scale (GRS) scores. The study attempted to answer the following research questions: 1) Will there be a significant difference in In-Service Board exams and Global Rating Scales scores, between the historical control group and the Simulation Wars intervention group 2) Will there be trends in resident performance in those who participated in Simulation Wars with those residents attending more Simulation Wars having improved Global Rating Scores and In-Service Board scores 3) Will the intervention groups' involvement in a specific subcategory topic of Simulation Wars affect their In-service Board score for that subcategory topic.

The remainder of this chapter describes the results obtained in the study and will include the following sections: (1) quantitative results for the first research question; (2) quantitative results for the second research question; (3) quantitative results for the third research question.

The study cohort contained 70 participants who attended Simulation Wars and 57 controls. The 70 participants who attended Simulation Wars (intervention group) were first, second and third year EM residents training at a teaching hospital in a large city in the Southwestern United States between 2013- 2015 academic years. The historical control (control group) consisted of EM residents who were trained between 2009-2012, but who did not participate in Simulation Wars.

Descriptive Statistics

The control and intervention groups were not of equal numbers. This was due to a different number of residents accepted every year. So in more recent years the number of residents was close to 30, whereas three years before it was close to 20. The researcher only requested three years of control subjects and three years of intervention subjects. Table 1 represents descriptive statistics.

Table 1. Descriptive Statistics

Variable Name	**Total N(%)**	**Cases N(%)**	**Controls N(%)**
	127 (100)	70 (55.1)	57 (44.9)
Age, yrs mean (median)	28.4 (27)	28.4 (27)	28.5 (27)
Sex			
Women	49 (38.6)	23 (32.9)	26 (45.6)
Men	78 (61.4)	47 (67.1)	31 (54.4)
Race			
White	79 (62.2)	41 (58.6)	38 (66.7)
Non White	48 (37.8)	29 (41.4)	19 (33.3)
Degree			
MD	112 (88.2)	64 (91.4)	48 (84.2)
DO	15 (11.8)	6 (8.6)	9 (15.8)
Medical School			
US	122 (97.6)	68 (100)	54 (94.7)
Non US	3 (2.4)	0	3 (5.3)
Scores			
YR1Score, mean (range)	71.6 (54 – 87)	71.7 (54 – 87)	71.6 (60 – 83)
YR1Percentage rank(range)	.53 (.05 – 98)	.56 (.05 – 98)	.51 (.05 – 90)
YR 2 Score mean (range)	77.4 (57-90)	76.9 (57-90)	78.5 (66-89)
YR 2 Percentage rank(range)	.56 (.02-.98)	.561 (.02-.98)	.57(.05-96)
YR 3 Score, mean (range)	77.9 (94-100)	78.7 (62-100)	77.4 (84-92)
YR3 Percentage rank (range)	.49 (.02-97)	.49 (.02-.85)	.50 (.02-97)
GRSPC			
AE	1 (0.8)	1 (1.4)	0
ME	118 (97.5)	67 (97.1)	51 (98.1)
BE	2 (1.7)	1 (1.4)	1 (1.9)
GRSMK			
AE	2 (1.6)	2 (2.9)	0
ME	117 (95.9)	66 (95.7)	51 (96.2)
BE	3 (2.5)	1 (1.4)	2 (3.8)
GRSPBP			
AE	1 (0.8)	0	1 (1.9)
ME	118 (96.7)	68 (98.6)	50 (94.3)
BE	3 (2.5)	1 (1.4)	2 (3.8)
GRSCOMM			

AE	10 (8.2)	0	10 (18.9)
ME	109 (89.3)	66 (95.7)	43 (81.1)
BE	3 (2.5)	3 (4.3)	0
GRSPROF			
AE	12 (9.8)	0	12 (22.6)
ME	108 (88.5)	67 (97.1)	41 (77.4)
BE	2 (1.6)	2 (2.9)	0
GRSSBP			
AE	0	0	0
ME	121 (99.2)	69 (100)	52 (98.1)
BE	1 (0.8)	0	1 (1.9)

N=Number, YR=Year, MD=Medical Doctor, DO=Doctor of Osteopthic Medicine,
US=US Medical School, nonUS=Non US Medical School
YR1=First year residents, YR2=Second year residents, YR3=Third year residents
Score=In-service score, Percentage Rank= percent of all resident in ACGME accredited EM residency programs, according to training level, who scored the same or lower than the resident of that year's exam.
GRS=Global Rating Score, PC=Patient Care, MK=Medical Knowledge, PBP=Practice Based Performance, Comm=Communication & Interpersonal Skills, Prof=Professionalism, SBL=System Based Practices,
AE=Above Expectations, ME=Met Expectations, BE=Below Expectations

The control and intervention group (cases) included similar groups according to gender (Figure 1), race (Figure 2), degree (Figure 3) and type of medical school attended.

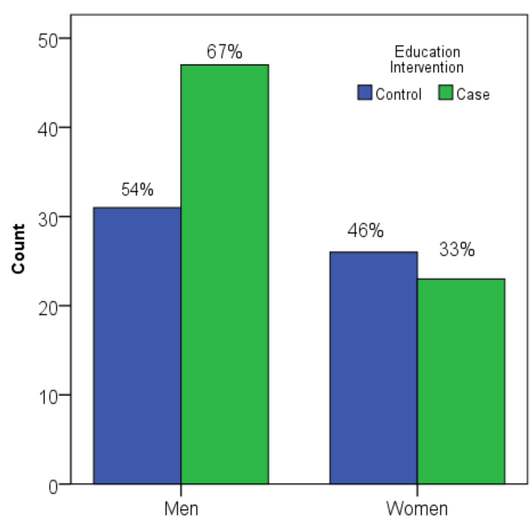

Figure 1. Distribution of cases and controls by gender was not significantly different.

.

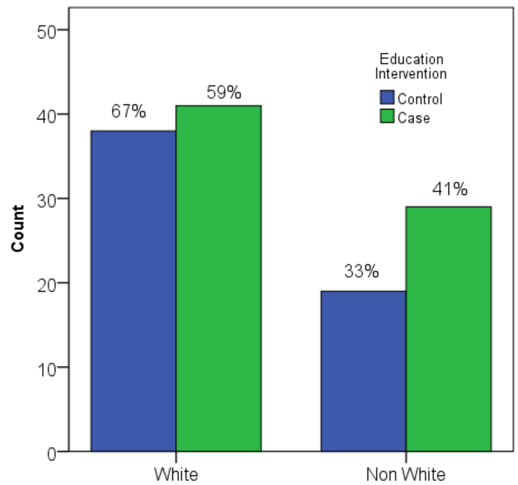

Figure 2. Distribution of cases and control by race was not significant.

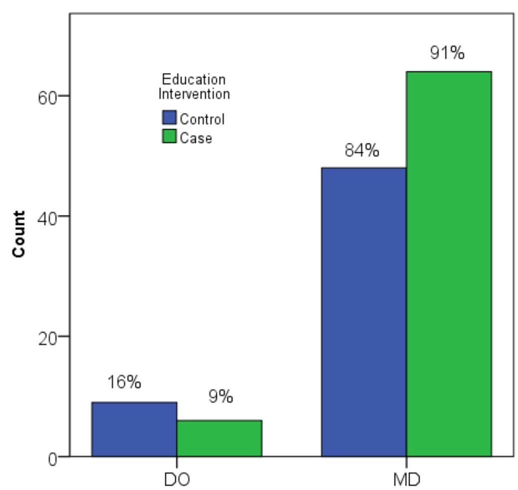

Figure 3. Distribution of cases and controls by professional background; Medical Doctor (MD) vs. Doctor of Osteopathic Medicine (DO) was not significant.

Results for the First Question

The first research question asked: Is there a difference in the In-Service Board exams and Global Rating Scale scores, between the historical control group and the Simulation Wars intervention group?

On the analyses of differences between cases and controls, nominal variables were analyzed by using the Chi-2 test or Fisher's exact test. Continuous variables were analyzed with the t-test if normally distributed or parametric tests such as the Mann-Whitney if a non-normal distribution was found.

Table 2. In-Service Board Exams and Global Rating Scale Scores, between Historical Control Group and Simulation Games Intervention Group (Cases)

Variable Name		**Cases N(%)**	**Controls N(%)**	**p-value**
		70 (55.1)	57 (44.9)	
YR1Score, mean (range)		71.7 (54 – 87)	71.6 (60 – 83)	.99
YR1Percentage, mean (range)		.56 (.05 – 98)	.51 (.05 – 90)	.41
YR 2 Score mean (range)		76.9 (57-90)	78.5 (66-89)	.24
YR 2 Percentage, mean(range)		.561 (.02-.98)	.57(.05-96)	.74
YR 3 Score, mean (range)		78.7 (62-100)	77.4 (84-92)	.55
YR3 Percentage, mean(range)		.49 (.02-.85)	.50 (.02-97)	.87
GR1PC				.67
AE		1 (1.4)	0	
ME		67 (97.1)	51 (98.1)	
BE		1 (1.4)	1 (1.9)	
GR1MK				.33
AE		2 (2.9)	0	
ME		66 (95.7)	51 (96.2)	
BE		1 (1.4)	2 (3.8)	
GR1PBP				.37
AE		0	1 (1.9)	
ME		68 (98.6)	50 (94.3)	
BE		1 (1.4)	2 (3.8)	
GR1COMM				<.001
AE		0	10 (18.9)	
ME		66 (95.7)	43 (81.1)	
BE		3 (4.3)	0	
GR1PROFE				<.001
AE		0	12 (22.6)	
ME		67 (97.1)	41 (77.4)	
BE		2 (2.9)	0	
GR1SYST				.43**
ME		69 (100)	52 (98.1)	
BE		0	1 (1.9)	

There were no significant differences in In-Service Board exams scores between the historical control group and the Simulation Games intervention group/(cases (Table 2) but there was a trend for improvement (Figure 4).

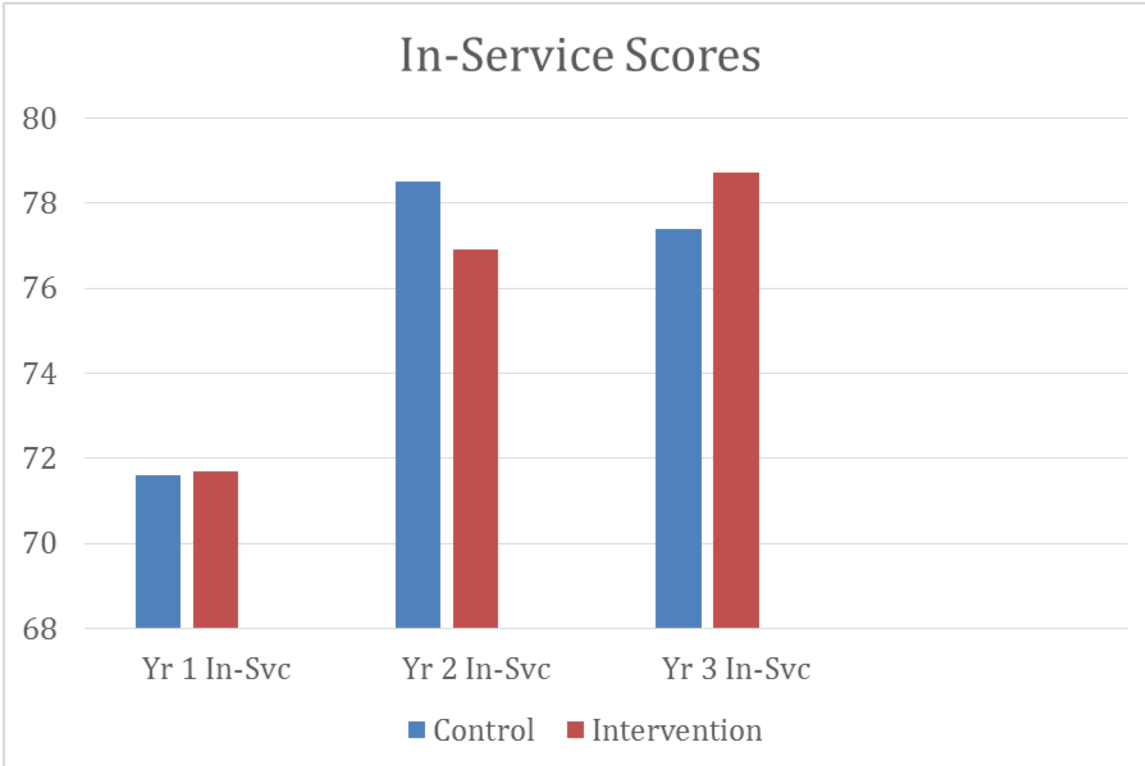

Figure 4. In-service scores, trend for improvement from YR1 to YR 2 to YR3.

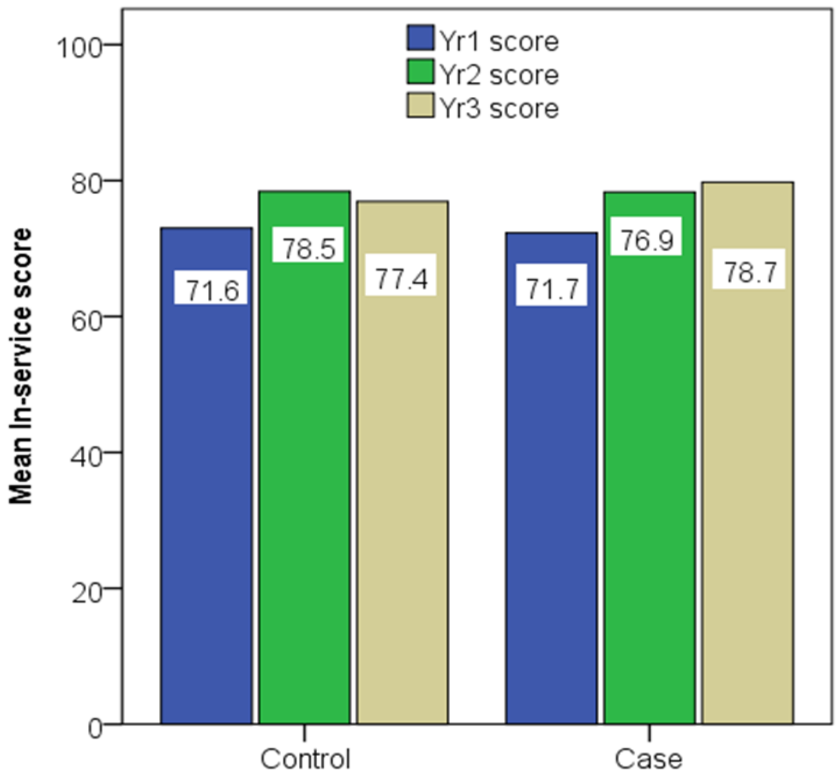

Figure 5. In-Service scores by year in residency.

.

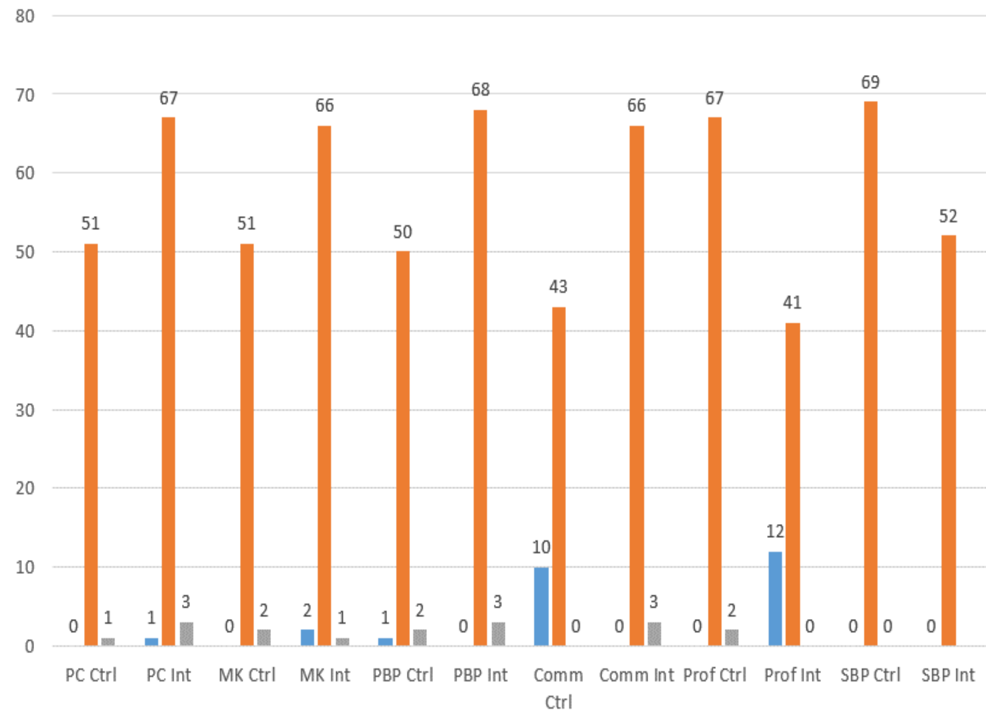

Figure 6. Global rating scores for ACGME competencies.

(AE = Above Expectations) (ME = Met Expectations) (BE = Below Expectations)

(Ctrl = Control) (Int = Intervention) (PC = Patient Care) (MK = Medical Knowledge)

(PBL = Practice Based Performance) (Com = Communication) (Prof = Professionalism)

(SBP = System Based Practices)

.

There was a significant difference in Global Rating Scale scores only for ACGME competencies of communication ($p<.001$) and professionalism ($p<.001$) which were significantly better for the control group, which had a higher number of AE scores than the intervention group (Figure 6). For all other competencies there were no significant difference in GRS scores. This held true also for GRS scores at each level of training (Figure 5 & 6) (Table 3).

Table 3. GRS Scores by Year in Residency

Variable Name	Total N(%)	Cases N(%)	Controls N(%)	p-value
	127 (100)	70 (55.1)	57 (44.9)	
GR1PC				.67
AE	1 (0.8)	1 (1.4)	0	
ME	118 (97.5)	67 (97.1)	51 (98.1)	
BE	2 (1.7)	1 (1.4)	1 (1.9)	
GR1MK				.33
AE	2 (1.6)	2 (2.9)	0	
ME	117 (95.9)	66 (95.7)	51 (96.2)	
BE	3 (2.5)	1 (1.4)	2 (3.8)	
GR1PBP				.37
AE	1 (0.8)	0	1 (1.9)	
ME	118 (96.7)	68 (98.6)	50 (94.3)	
BE	3 (2.5)	1 (1.4)	2 (3.8)	
GR1COMM				<.001
AE	10 (8.2)	0	10 (18.9)	
ME	109 (89.3)	66 (95.7)	43 (81.1)	
BE	3 (2.5)	3 (4.3)	0	
GR1PROFE				<.001
AE	12 (9.8)	0	12 (22.6)	
ME	108 (88.5)	67 (97.1)	41 (77.4)	
BE	2 (1.6)	2 (2.9)	0	
GR1SYST				.43**
ME	121 (99.2)	69 (100)	52 (98.1)	
BE	1 (0.8)	0	1 (1.9)	
GR2PC				.04
AE	8 (7.2)	1 (1.7)	7 (13.2)	
ME	102 (91.9)	56 (96.6)	46 (86.8)	
BE	1 (0.9)	1 (1.7)	0	
GR2MK				.18
AE	8 (7.3)	2 (3.5)	6 (11.3)	
ME	101 (91.8)	54 (94.7)	47 (88.7)	
BE	1 (0.9)	1 (1.8)	0	
GR2PBP				.99**
AE	2 (1.8)	1 (1.8)	1 (1.9)	
ME	108 (98.2)	56 (98.2)	52 (98.1)	
GR2COMM				.002**
AE	12 (10.9)	1 (1.8)	11 (20.8)	
ME	98 (89.1)	56 (98.2)	42 (79.2)	
GR2PROFE				.003**
AE	11 (10)	1 (1.8)	10 (18.9)	
ME	99 (90)	56 (98.2)	43 (81.1)	
GR2SYST				.35
AE	4 (3.6)	1 (1.8)	3 (5.7)	
BE	106 (96.4)	56 (98.2)	50 (94.3)	
GR3PC				.006**
AE	14 (16.1)	1 (2.9)	13 (25)	
ME	73 (83.9)	34 (97.1)	39 (75)	

GR3MK				.01**
AE	9 (10.2)	0	9 (17)	
ME	79 (89.8)	35 (100)	44 (83)	
GR3PBP				.04**
AE	7 (8)	0	7 (13.2)	
ME	81 (92)	35 (100)	46 (86.8)	
GR3COMM				.003**
AE	11 (12.5)	0	11 (20.8)	
ME	77 (87.5)	35 (100)	42 (79.2)	
GR3PROFE				.05
AE	10 (11.4)	1 (2.9)	9 (17)	
ME	78 (66.6)	34 (97.1)	44 (83)	
GR3SYST				.14
AE	8 (9.1)	1 (2.9)	7 (13.2)	
BE	80 (90.9)	34 (97.1)	46 (86.8)	

**Fisher's exact test.

The competencies of Communication and Professionalism were significantly different for all three years of residency with higher evaluations for the controls. Residents (controls) in their second year had higher evaluations in Patient Care. Residents (controls) in their third year had higher evaluations in Medical Knowledge, and Practice Based Performance

Results for Second Question

The second research question asked: Is there a trend in resident performance in the group that participated in Simulation Games? The t-test was used to compare subcategories of scores of the participants who did and did not attend a specific Simulation Wars session.

In the next table the difference were compared between In-Service scores by year, controlling for attendance to education intervention. For this analysis a t- test was used. Because some of the variables measuring differences were not normally distributed, parametric tests, (specifically the Mann Whitney U) were performed; the findings were similar.

Table 4. In-Service Scores for Intervention Group and Controls

Variable Name	**Total**	**Intervention**	**Controls**	**p-value**
Score differences				
Yr2 minus Yr1 score, mean (sd)	5.13 (6.66)	5.15 (7.27)	5.08 (5.73)	.96
Yr3 minus Yr1 score, mean (sd)	6.22 (12.1)	7.47 (9.01)	5.75 (13.20)	.62
Yr3 minus Yr2 score, mean (sd)	-.07 (11.95)	1.23 (7.04)	-1.35 (15.31)	.37
Percentage score differences				
Yr2 percent score minus Yr1 percent score, mean (sd)	.0094 (.22)	.02 (.20)	-.01 (.24)	.49
Yr3 percent score minus Yr1 percent score, mean (sd)	-.33 (.25)	-.09 (.19)	-.01 (.27)	.27
Yr3 percent score minus Yr2 percent score, mean (sd)	-.05 (.23)	-.05 (.22)	-.06 (.25)	.86

To understand if a substantial increment of the In-Service scores over the previous year were more likely observed in cases than in controls, we compared the difference in In-Service scores by year, controlling for attendance to the Simulation Wars intervention (Table 4). For this analysis a t-test was used. Because some of the variables measuring differences were not normally distributed, I performed parametric tests (specifically the Mann Whitney U) and the findings were similar. The largest difference observed was between year three and year one, as expected given the larger exposure to the Simulation Wars program of the third year residents compared to the first year residents. Overall, there were no significant differences for the Intervention group and Controls for In-Service scores by year of residency (Table 4).

The next analysis included comparisons of the Intervention groups' In-Service scores. The scores according to year in residency were compared.

Table 5. Comparisons of Score Assessments at Baseline (YR1) and at YR2 and YR3 of Intervention Group

Score Before N Mean (sd)	Score After N Mean (sd)	Mean diff	95% CI	T test	p-value
YR1Score N=52 71.65 (7.99)	YR2Score N=52 76.81 (7.47)	-5.15	-9.17, - 3.13	T(51)=-5.11	≤.001
Yr1Score N=17 72.29 (8.44)	YR3Score N=17 79.76 (8.28)	-7.47	-12.11, -2.83	T(16)=3.42	.004
YR2Score N=34 77.71 (6.73)	YR3Score N=34 78.9 (7.12)	-1.24	-3.69, 1.22	T(33)=1.02	.31

Note. The score shown in column 1 will always be the reference value. For this analysis, a matched paired *t* test was used. The N= indicates the pairs included, i.e., 52 when comparing year 1 with year 2, etc.

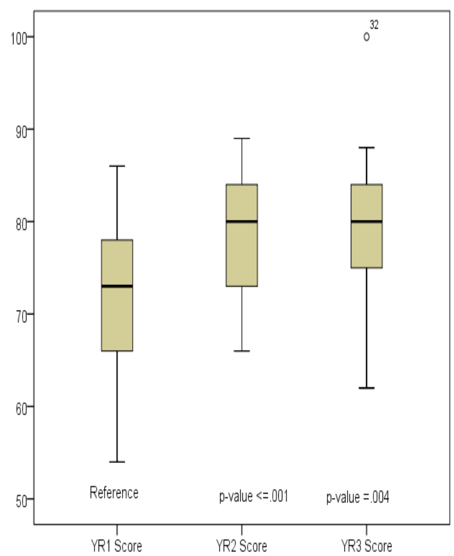

Figure 7. Comparison of in-service scores at YR1, YR2, and YR3 of the Intervention Group.

There was a significant improvement from YR1 to YR2 to YR3 for the Intervention Group (Table 5) (Figure 8).

Table 6. Comparisons of Score Assessments at Baseline (YR1) and at YR2 and YR3 of the Controls

Score Before N Mean (sd)	Score After N Mean (sd)	Mean diff	95% CI	T test	p-value
YR1Score N=35 73.20 (5.16)	YR2Score N=35 78.29 (5.62)	-5.09	-7.05, - 3.12	T(34)=-5.25	≤.001
Yr1Score N=45 71.51 (5.57)	YR3Score N=45 77.26 (12.77)	-5.75	-9.72, -1.78	T(44)=-2.91	.005
YR2Score N=35 78.46 (5.46)	YR3Score N=35 77.11 (14.18)	-1.35	-3.91, 6.60	T(34)=.52	.61

Note. Score shown in column 1 will always be the reference value. For these analyses, a matched paired *t* test was used. The N= indicates the pairs included, i.e., 35 when comparing YR1 with YR2, etc.

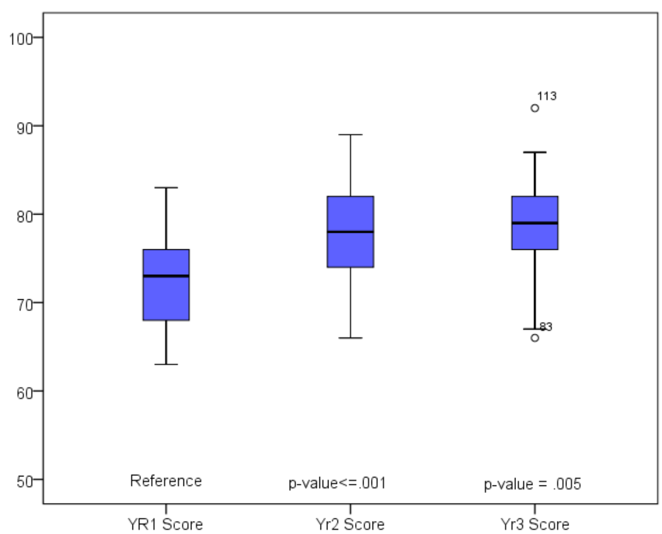

Figure 8. Comparison of scores at YR1, YR2, and YR3 of Controls.

There was a significant improvement from YR1 to YR2 to YR3 for the Control Group as well as the experimental group (Table 6) (Figure 8).

Results for Third Question

The third research question asked: Is there a difference in the intervention groups' In-service Board subcategory topic score according to whether they participated in a

Simulation War that matched that subcategory topic? The score showed in column 1 was always the reference value. For these analyses, a matched paired t test was used. The N= indicates the pairs included i.e. 45 comparing year 1 with year 2, etc. The mean value represents the mean of the percentage score. The majority In-Service board scores for those who were involved in Simulation Wars did not significantly differ if they were involved in Year 1, Year 2 or Year 3 compared to those who were not involved in Simulation Wars.

Table 7. Subcategories of In-Service Score by Attendance to Sim Wars Sessions on Same Topic in First-Year Residents

Subcategory	All Mean (sd)	No attendance to SIM session on topic of subcategory		Attendance to one or more SIM sessions on topic of subcategory		p-value
		N	Mean (sd)	N	Mean (sd)	
SSP1percent	.7079 (.1063)	7	.7557 (.0680)	45	.7004 (.1098)	.20
Abgas1percent	.7354 (.1479)	6	.5900 (.7000)	46	.7543 (.1215)	.17
Cardio1percent	.7083 (.1395)	3	.7833 (.1499)	49	.7037 (.1392)	.34
Enviro1percent	.6938 (.2097)	3	.7100 (0)	49	.6929 (.1261)	.58
Musc1percent	.7848 (.1670)	29	.7772 (.1574)	23	.7943 (.1815)	.72
OB1percent	.6892 (.1927)	3	.7200 (.0557)	49	.6873 (.1982)	.78
Psycho1Bpercent	.6462 (.2023)	50	.6440 (.2051)	2	.7000 (.1414)	.71
RenalUro1percent	.6900 (.1842)	51	.6937 (.1840)	1	.5000 (NC)	.50
Thoracic1percent	.7052 (.1201)	18	.6539 (.1419)	34	.7324 (.0986)	.02
Toxic1percent	.7583 (.1598)	3	.6667 (.1930)	49	.7639 (.1581)	.31
Trauma1percent	.7088 (.1234)	10	.6850 (.1692)	42	.7145 (.1117)	.61
Nerv1percent	.7713 (.1593)	22	.7336 (.1670)	30	.7990 (.1329)	.12

There was a significant ($p<.05$) difference the In-Service Board score in the subcategory of Thoracic Disorders in those first year residents who participated in the Thoracic Simulation Wars. All other types of Simulation Wars did not have a significant difference in subcategory In-Service Board scores (Table 7).

Table 8. Subcategories of In-Service Score by Attendance to Simulation Wars Sessions on Same Topic in Second-Year Residents

Subcategory	All Mean (sd)	No attendance to SIM session on topic of subcategory		Attendance to 1 or more SIM sessions on topic of subcategory		p-value
		N	Mean (sd)	N	Mean (sd)	
SSP1percent	.7512 (.1187)	24	.7721 (.1251)	45	.7400 (.1149)	.29
Abgas1percent	.7570 (.1092)	10	.6640 (.1157)	59	.7610 (.1026)	.008
Cardio1percent	.7509 (.1136)	20	.7635 (.1092)	49	.7457 (.1161)	.56
Enviro1percent	.7654 (.2013)	7	.8071 (.1345)	62	.7606 (.2078)	.44
Musc1percent	.7880 (.1596)	45	.7833 (.1618)	24	.7977 (.1584)	.74
OB1percent	.7714 (.1366)	11	.7718 (.1508)	58	.7714 (.1352)	.99
Psycho1Bpercent	.7709 (.1810)	67	.7210 (.1819)	2	.7250 (.2051)	.96
RenalUro1percent	.6984 (.1597)	68	.7003 (.1601)	1	.5700 (NC)	.42
Thoracic1percent	.7475 (.1236)	34	.7744 (.1321)	35	.7214 (.1081)	.07
Toxic1percent	.7735 (.1575)	7	.6829 (.1668)	62	.7837 (.1545)	.11
Trauma1percent	.7865 (.1116)	27	.7878 (.1327)	42	.7857 (.0974)	.95
Nerv1percent	.7974 (.1601)	39	.7882 (.1635)	30	.8093 (.1588)	.59

Note. SD, standard deviation; NC, not calculated.

There was a significant difference in the Abdominal and GI In-Service Board scores for those second year residents who attended the Abdominal and GI Simulation Games (Table 8).

Table 9. Subcategories of In-Service Score by Attendance to Sim Wars Sessions on Same Topic in Third-Year Residents

Subcategory	All Mean (sd)	No attendance to SIM session on topic of subcategory		Attendance to 1 or more SIM sessions on topic of subcategory		p-value
		N	Mean (sd)	N	Mean (sd)	
SSP1percent	.7866 (.0953)	22	.7718 (.0952)	13	.7846 (.0987)	.71
Abgas1percent	.7914 (.1068)	4	.7050 (.1250)	31	.8026 (.1012)	.09
Cardio1percent	.7180 (.1064)	19	.7132 (.1148)	16	.7238 (.0988)	.77
Enviro1percent	.7914 (.1602)	6	.7733 (.1727)	29	.7952 (.1604)	.77
Musc1percent	.8014 (.1581)	24	.7758 (.1712)	11	.8573 (.1123)	.11
OB1percent	.7276 (.1057)	9	.7333 (.1179)	26	.7242 (.1036)	.83
Psycho1Bpercent	.7100 (.2217)	34	.7100 (.2251)	1	.7100 (NC)	.99
RenalUro1percent	.6383 (.2150)	35	.6383 (.2150)		No sessions recorded	NA
Thoracic1percent	.7611 (.1194)	22	.7395 (.1167)	13	.7977 (.1194)	.17
Toxic1percent	.8166 (.1172)	4	.7775 (.1193)	31	.8216 (.1180)	.49
Trauma1percent	.8111 (.1098)	21	.7743 (.1095)	14	.8664 (.0873)	.01
Nerv1percent	.8349 (.1195)	25	.8376 (.1106)	10	.8280 (.1457)	.83

There was a significant ($p<.05$) difference in the In-Service Board score in the subcategory of Trauma Disorders for those third year residents who participated in the Trauma Simulation Wars. All other types of Simulation Wars did not have a significant difference in subcategory In-Service Board scores (Table 9).

Table 10. Paired Comparisons of Subcategories Score Assessments at Baseline (Year 1) and at Year 2 of the Intervention in Participants Who Attended One or More Simulation Wars on Same Topic

Subcategory	1st year		2nd year		p-value
	N	Mean (sd)	N	Mean (sd)	
SSPpercent	45	.700 (.110)	45	.740 (.115)	.10
Abgaspercent	46	.754 (.122)	59	.761 (.103)	.76
Cardiopercent	49	.704 (139)	49	.746 (.116)	.11
Endopercent	23	.732 (.200)	36	.774 (.219)	.45
Enviropercent	49	.693 (.216)	62	.761 (.208)	.10
Headpercent	48	.704 (.187)	60	.750 (.159)	.17
Muscpercent	23	.794 (.182)	24	.797 (.158)	.96
OBpercent	49	.687 (.198)	58	.771 (.135)	.01
PsychoBpercent	2	.700 (.1414)	2	.715 (.2051)	NC
RenalUropercent	1	.50 (NA)	1	.57 (NA)	NC
Thoracicpercent	34	.732 (.099)	35	.721 (.108)	.66
Toxicpercent	49	.764 (.158)	62	.784 (.155)	.51
Traumapercent	42	.715 (.112)	42	.786 (.097)	.003
Nervpercent	30	.799 (.133)	30	.809 (.159)	.79

*Note.*The subscores for Psycho Behavior and RenalUro have few participants, comparisons are not possible. NA=Not available, NC=Not calculated.

There was a significance (p<.05) difference in the In-Service Board scores in the subcategory of Trauma Disorders and ObGyn in those who participated in a matched topic of Simulation Wars. All other types of Simulation Wars did not have a significant difference in subcategory In-Service Board scores (Table 10).

ObGyn and Trauma were the only subcategories of In-Service scores where there was a significant difference from first year to third year students if they attended a Simulation Wars on ObGyn or trauma (Table 11).

Table 11. Paired Comparisons of Subcategory Score Assessments at Baseline (Year 1) and at Year 3 in Intervention (Cases) Who Attended Sessions in Related Topic

Score Before N Mean (sd)	Score After N Mean (sd)	Mean diff	95% CI	T test	p-value
Signs & Symptoms Yr1percent rank score N=45 .7004 (.1098)	Signs & Symptoms Yr2percent rank score N=45 .7400 (.1149)	-.0396	-.0820, .0029	T(44)=-1.877	.07
Abd&GI Yr1percent rank score N=45 .7260 (.1558)	Abd&GI Yr2percent rank score N=45 .7613 (.1203)	-.0353	-.1090, .0192	T(44)=-1.305	.20
Cardiology Yr1percent rank score N= 45 .7013 (.1436)	Cardiology Yr2percent rank score N= 45 .7416 (.1180)	-.0402	-.0916, .0111	T(44)=-1.579	.12
Environmental Yr1percent rank score N=45 .6944 (.2142)	Environmental Yr2percent rank score N=45 .7569 (.2030)	-.0624	-.1419, .0170	T(44)=-1.583	.12
Immunological Yr1percent rank score N=45 .6831 (2508)	Immunological Yr2percent rank score N=45 .6513 (.2141)	.0317	-.0641,0127	T(44)=.668	.507
Systemic Infections Yr1percent rank score N=45 .7604 (.1363)	Systemic Infections Yr2percent rank score N=45 .7409 (.1599)	.0196	-.0468,.0859	T(44)=.593	.566
Musculoskeletal Yr1percent rank score N=45 .7787 (.1597)	Musculoskeletal Yr2percent rank score N=45 .7836 (.1605)	-.0049	-.0662,.0585	T(44)=-.161	.873
Ob/Gyn Yr1percent rank score	Ob/Gyn Yr1percent rank score	-.0864	.0338,-.1546	T(44)=-2.55	.014

N=45 .6851 (.19112)	N=45 .7716 (.1428)				
Psycholog/Behavior Yr1percent rank score N=45 .6562 (.2035)	Psycholog/Behavior Yr2percent rank score N=45 .7213 (.1758)	-.0651	.0410,- .1478	T(44)=- 1.58	.120
Thoracic Yr1percent rank score N=45 .7076 (.1054)	Thoracic Yr2percent rank score N=45 .7213 (.1182)	-.0137	.02479,- .0637	T(44)=- .556	.581
Toxicology Yr1percent rank score N=45 .7589 (.1620)	Toxicology Yr2percent rank score N=45 .7864 (.1388)	-.0275	.0265,- .0811	T(44)=- .1.03	.305
Trauma Yr1percent rank score N=45 .7149 (.1185)	Trauma Yr2percent rank score N=45 .7822 (.1056)	-.0673	.0232,- .1141	T(44)=- 2.90	.006
Nervous System Yr1percent rank score N=45 .7724 (.1489)	Nervous System Yr2percent rank score N=45 .7936 (.1543)	-.0211	.0289,- .0795	T(44)=- .728	.470

Thoracic and Trauma were the only subcategories of In-Service scores where there was a signiciant difference from second year to third year if they attended a Simulation Was on Thoracic or Trauma (Table 12).

Table 12. Paired Comparisons of Subcategories Score Assessmentsa Year 2 and Year 3 of the Intervention in Participants Who Attended Sessions in Related Topic

Score in year 2 N Mean (sd)	Score in year 3 N Mean (sd)	Mean diff	95% CI	T test	p-value
Signs & Symptoms Yr2percent rank score N= 13 .7200 (.1225)	Signs & Symptoms Yr3percent rank score N= 13 .7846 (.0987)	-.0646	-.1589, .0297	T(12)= -1.49	.16
Abd&GI Yr2percent rank score N= 30 .7480 (.0924)	Abd&GI Yr3percent rank score N= 30 .8000 (.1019)	-.0520	-.0939, -.0101	T(29)= -2.54	.07
Cardiology Yr2percent rank score N= 16 .7556 (.1027)	Cardiology Yr3percent rank score N= 16 .7238 (.0988)	.0319	-.0453, .1091	T(15)= .88	.39
Environmental Yr2percent rank score N= 29 .7872 (.2387)	Environmental Yr3percent rank score N= 29 .7952 (.1604)	-.0079	-.1110, .0952	T(28)= -.158	.88
HEENT Yr2percentage rank score N= 34 .7985 (.1505)	HEENT Yr3percentage rank score N= 34 .7659 (.1333)	.0326	-.0312, .0965	T(33)= 1.040	.31
Immunological Yr2percent rank score N=	Immunological Yr3percent rank score N=				
Systemic Infections Yr2percent rank score N= 34 .7562 (.1641)	Systemic Infections Yr3percent rank score N= 34 .7588 (.1577)	-.0026	-.0723, .0670	T(33)= -.077	.94
Musculoskeletal Yr2percent rank score N= 11 .8300 (.1270)	Musculoskeletal Yr3percent rank score N= 11	-.0273	-.1228, .0683	T(10)= -.636	.54

	.8573 (.1123)				
Ob/Gyn Yr2percent rank score N= 25 .7660 (.1416)	Ob/Gyn Yr3percent rank score N= 25 .7248 (.1057)	.0412	-.0254, .1078	T(24)= 1.276	.21
Psycholog/Behavior Yr2percent rank score N= 1 .8600	Psycholog/Behavior Yr3percent rank score N= 1 .7100				NC
Thoracic Yr2percent rank score N= 13 .6731 (.1069)	Thoracic Yr3percent rank score N= 13 .7977 (.1194)	-.1246	-.2142, -.0350	T(12)= -3.30	.010
Toxicology Yr2percent rank score N= 30 .7650 (.1847)	Toxicology Yr3percent rank score N= 30 .8240 (.1192)	-.0590	-.1409, .0229	T(29)= -1.47	.15
Trauma Yr2percent rank score N= 14 .7657 (.1155)	Trauma Yr3percent rank score N= 14 .8664 (.0873)	-.1007	-.1698,-.0317	T(13)= -3.15	.008
Nervous System Yr2percent rank score N= 10 .7520 (.2264)	Nervous System Yr3percent rank score N= 10 .8280 (.1457)	-.0760	-.2090, .0570	T(9)= -1.29	.23

Figures for relevant associations comparing subcategory scores by year and attendance to the particular session of Abdominal GI YR1 vs YR3

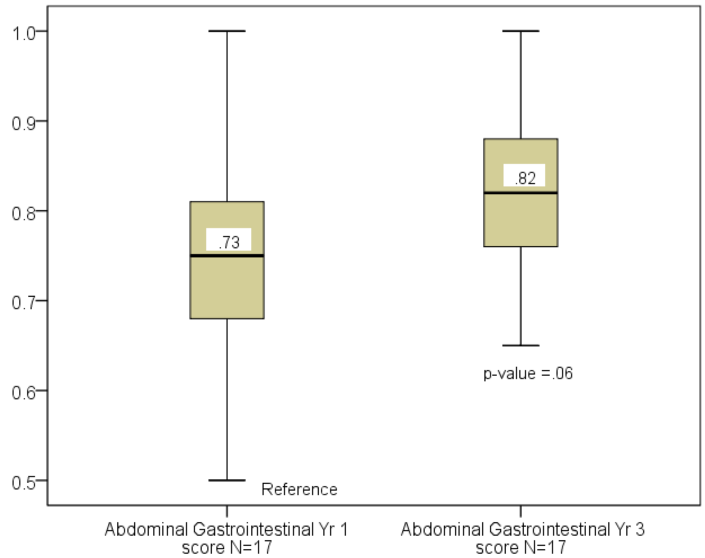

Figure 9. Significant subcategories on In-Service test: Session Abdominal GI YR1 vs. YR3.

Session Trauma YR1 vs. YR3 Note on the outlier, the diff (YR3 score minus YR1 score) is normally distributed, the assumptions needed to perform the paired analyses are met)

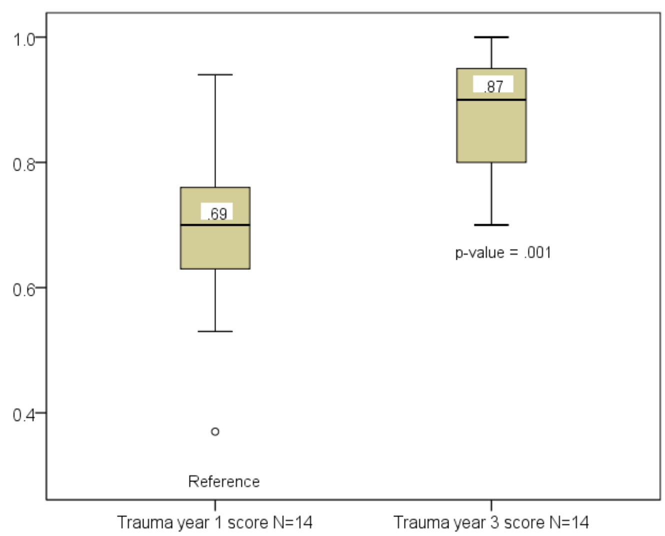

Figure 10. Significant subcategories on In-Service test: Session Trauma YR1 vs. YR3.

Session Trauma YR1 vs. YR 2 Note on the outlier, the diff (YR2 score minus YR1 score) is normally distributed, the assumptions needed to perform the paired analyses are met)

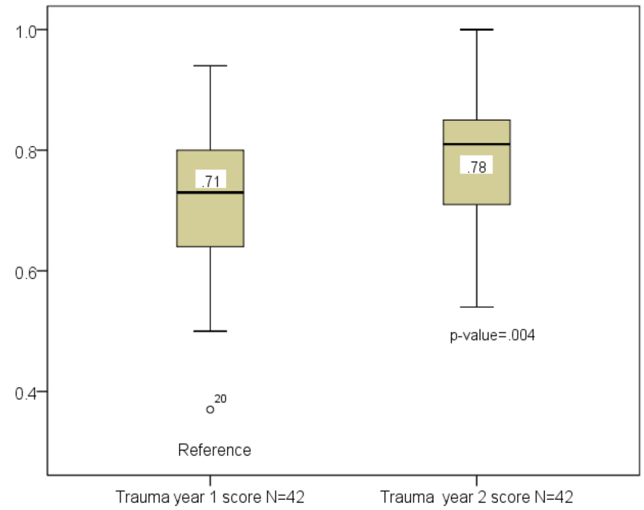

Figure 11. Significant subcategories on In-Service test: Session Trauma YR2 vs. YR3.

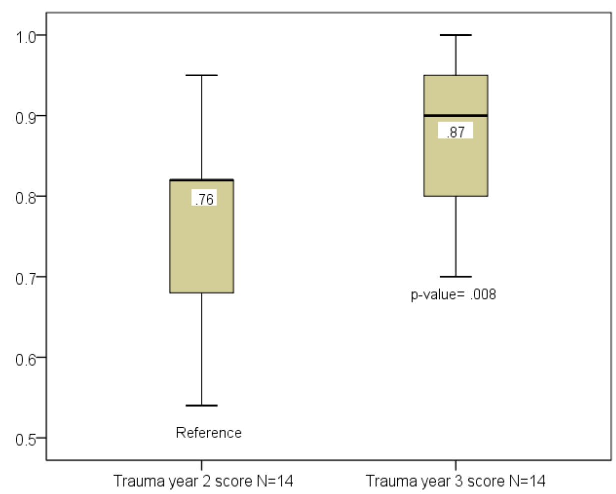

Figure 12. Significant subcategories on In-Service test: Session OBGYN YR1 vs. YR2.

Session OBGYN YR 1 vs. YR 2 Note on the outliers, the diff (YR2 score minus YR 1 score) is normally distributed, the assumptions needed to perform the paired analyses are met

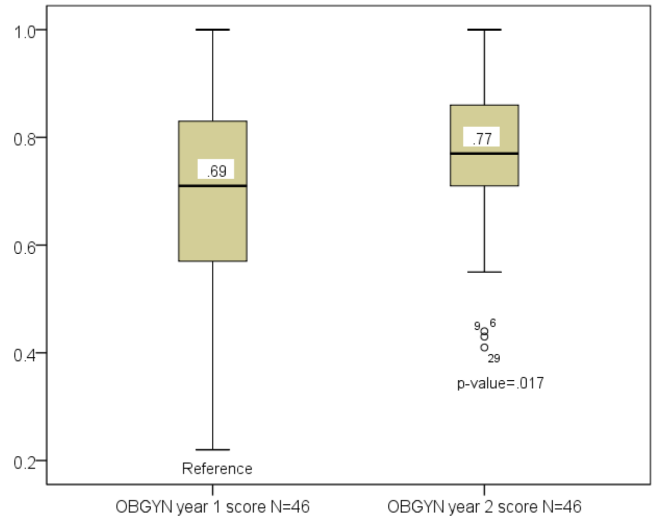

Figure 13. Significant subcategories on In-Service test: Session Thoracic YR1 vs. YR2.

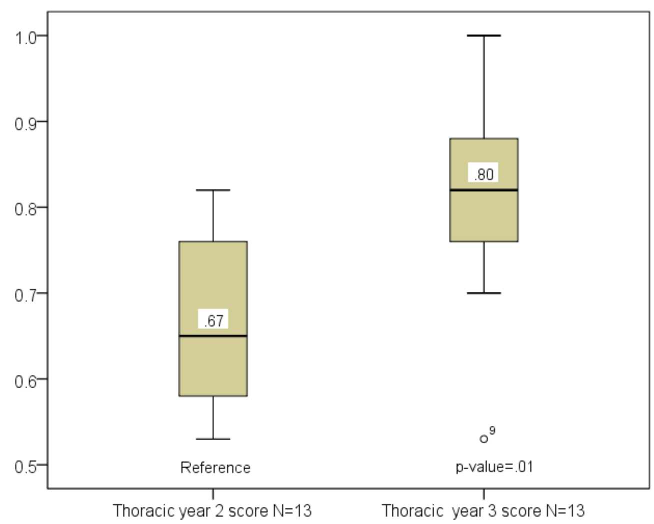

Figure 14. Significant subcategories on In-Service test: Session Thoracic YR2 vs. YR3.

Chapter V

Discussions and Conclusions

There are several methods that can be utilized to measure clinical reasoning skills. Medical residents are typically evaluated with classic written examinations, GRS scores, patient satisfaction surveys and objective structured clinical examinations and simulation. The GRS scores are the most frequently used by Graduate Medical Education (GME) residency programs. The GRS scores evaluate the resident on variables that include medical knowledge and behaviors by the faculty that observed them while working in the ED. The six ACGME competencies for which residents are graded per the GRS scores include; medical knowledge, patient care, practice based learning, system based practices, communication and professionalism (Littlefield, Paukert, & Schoolfield, 2001; Gray, 1996). GRS scores by faculty have been viewed as a better measure of the residents' ability to apply knowledge than other methods (Haurani, 2007). Cognitive knowledge is measured by test scores such as the In- Service Board scores and global rating scores but mainly with the In-Service Board score. Tests such as resident In-Service board scores measure more content knowledge with limited ability to measure how well the resident applies this knowledge to patients (Krishnamurthy, 2009). Our study included both the GRS scores and In-Service Board scores in order to include a measure of both medical knowledge and the application of this knowledge.

The In-Service board scores and GRS scores attempt to measure competency to ensure residents are more competent as they move from their first year to last year of residency. Competency in medicine is defined as well learned skills that, once acquired, may be used repeatedly on many patients. This should also involve the capacity to respond appropriately to multiple, simultaneous symptoms and medical events that can further complicate a patient's morbidity (Meterissain, et al., 2007).

In-Service board scores specifically in EM have been shown to be positively correlated to final certification board scores which evaluate clinical competency (Hamlin, Cutle, Hafner, &Thompson, 2008). Final certification board scores correlate with the Continuous Certification (ConCert) which is taken every 10 years by practicing emergency medicine physicians (Munger, 1982;Hamlin, 2008). The ConCert examination that measures clinical knowledge and how it is applied in medical practice and must be passed every 10 years to maintain certification. The ConCert examination measures diagnostic reasoning by using psychometrically validated items (Holmboe, Lipner, & Greiner, 2000). Final certification board scores were not available for this study. Since In-Service Board scores were available and since it has been shown they correlate with clinical competency as measured by the final board scores and Concert, In-Service Board scores served as a measure of clinical competency.

We evaluated if teaching clinical reasoning skills using a simulation method of Simulation Wars affect EM residents clinical reasoning skills as measured on In-Service Board exams and GRS scores. We chose to look at not just one level of residency training but first, second and third year residents who should impove their clinical reasoning through experience and whether Simulation Wars affected these skills.

First Research Question

The first research question asked: Is there a difference in the In-Service Board exams and GRS scores, between the historical control group and the Simulation Wars intervention group (cases)?

Surprisingly there were no differences between cases and controls for overall In-Service board scores for first, second and third year residents. Some studies have shown that there was improvement in In-Service board scores with clinical reasoning exercises. A clinical reasoning exercise, the surgical pattern recognition examination (PAT), was administered to surgical residents and medical students at one point in time during one academic year. The PAT was developed in 1988 and was found to be reliable and correlated to the National Board Medical Examination (Case & Swanson, 1989). The researchers found that their scores on the PAT were significantly higher in senior than junior residents (82% versus 63%, p=0.004) and significantly correlated to the Surgery In-Service Exam, r=.67 (Dunn & Woolliscroft, 1995). Perhaps our study results were different because unlike our study which evaluated analytic reasoning skills, the PAT evaluated clinical reasoning's nonanalytic skills. Our study focused on analytic clinical reasoning skills which develop over time with clinical experience and include higher intelligence (Durning, Dong, Artino, der Vleuten, Holmboe, & Schuwirth, 2014). Our residents were evaluated quarterly over a 2.5 year period versus one point in time as occurred in their study. These factors made our study more applicable to analytic clinical reasoning exercises.

In another study by Humbert and colleagues the script concordance test (SCT) was used to assess clinical reasoning in 37 emergency medicine clerkship and residency (Humbert, Besinger, & Miech, 2011). The SCT format starts with a short clinical vignette that is followed by a series of proposed diagnoses, investigational studies, or therapies that a clinician should know. The learner is then given one additional piece of information to the case and asked what the effect of that information would be on his or her clinical reasoning related to the diagnosis, test, or treatment. Test takers indicate their qualitative judgments for each item on a five-point Likert scale that ranges from -2 to +2. An example of this is the test taker being asked if:

- "… you are thinking of the following of appendicitis and you find a normal WBC count, the hypothesis becomes?
    - -2=highly unlikely,
    - -1=less likely than before
    - 0= neither more nor less likely,
    - +1=more likely than before,
    - +2=very likely.

The SCT-EM scores of the residents were linked with individual performance on the In-Service Board exam. Humbert and colleagues found a statistically significant correlation between SCT performance and in-training exam scores ($r=0.69$, $p<.001$). Unlike their study, our study did not use an SCT format and scoring of Simulation Wars. Our Simulation Wars format included a lot of the components of other validated and reliable clinical reasoning exercises such as the Script Concordance Test and Ottawa Crisis Resource Management Global Rating Scale. The Simulation exercises that have been implemented in residency programs as per the studies by Young in 2007 and Hedrick in 2010 also included a format that included that of validated and reliable methods. In the studies by Young, Hendrick and Humbert as well as ours, clinical reasoning's analytic process was emphasized. Humbert's study was over one academic year and the SCT was given once during that year. Since our study was over a longer time span and more frequent than theirs perhaps this was another reason for our results being different. Consequently, these results possibly present a better reflection of a clinical reasoning exercise affecting In-Service Board Scores.

Another study of surgical residents (30) undergoing simulation as part of a boot camp at the beginning of their first year of residency showed a significant correlation with their in-service exam (Fernandez et al, 2012). The patient cases included shock, surgical emergencies and respiratory, cardiac and trauma management. They concentrated on analytic processes like ours but their clinical reasoning exercise exam was only done at one point in time over one academic year. In addition to our larger sample size and longer period of study, we included a larger selection of cases that were specific to EM, including Signs and Symptoms, Abdominal/Gastrointestinal, Cardiology, Environmental, Head Eyes Ears Nose and Throat, Thoracic, Toxiciology and Nervous System. Perhaps their study results were not consistent with ours since they did not include a control group. Since it is known that In-Service scores increase from first year to second year and from second year to third year of residency, the inclusion of a control group ensured that the intervention of Simulation Wars did affect In-Service board scores. Both our intervention group and control group showed a significant improvement in In-Service board scores from first year to third year of residency, showing that the Simulation Wars didn’t have an overall effect on In-Service board scores.

In our study there was not a significant statistical difference in the majority of GRS scores for those in the intervention group and control group in all three years of residency. There was a significant difference, however, between GRS scores for the competencies of Communication and Professionalism in the control group and not rhw intervention group. Surprisingly the control group had higher scores than the intervention group in Communication and Professionalism. This could have been due to inherent qualities such as the maturity or personality of those residents’ who scored higher. Maybe this finding was due to the historical control design with the faculty grading the control residents higher than the faculty who graded the intervention group in these areas of Communication and Professionalism. Another possible explanation was the "halo effect." The halo effect is a cognitive bias in which an evaluator’s overall impression of a person is influenced by the evaluator’s feelings and thoughts about the person’s character. The Halo Effect causes people to be biased in their judgments by transferring their feelings about one attribute to other, unrelated, attributes (Thorndike, 1920). One negative experience by the faculty evaluator with a resident may cause the faculty member to evaluate the resident negatively on all other aspects of that resident and in the future. Unfortunately, there were no other significant differences in the other competencies, such as Medical Knowledge or Patient Care. Since the competencies of Patient Care and Medical Knowledge are the competencies which would be thought to improve after the intervention of Simulation Wars since they include clinical reasoning. Medical Knowledge is evaluated as recommended by the ACGME by rating the resident on how he/she demonstrates appropriate medical knowledge in care of the ED patients. Patient Care is evaluated as recommended by the ACGM by rating how the resident basis management with all available data, narrows and prioritizes the list of differential diagnoses to determine appropriate management. We were surprised that our findings were not significantly improved in the competencies of Medical Knowledge and Patient Care with Simulation Wars.

Second Research Question

The second research question was: Is there a trend in resident performance in the group that participated in Simulation Wars in their In-Service scores? There was not a significant difference in overall In-Service board scores between the intervention group and controls. With paired t tests, however, there was a significant improvement in overall In-Service board scores from first year to second year but not from second year to third year for the controls or the intervention group. Similar to our results, in a study by Clarke and colleagues on EM resident crisis resource management ability with a simulation-based longitudinal study, they found a significant improvement from first year of residency to second year of residency but not from second year of residency to third year of residency (Clarke et al., 2014). Unlike our study the performance on the PAT a clinical reasoning test administered to surgical residents showed performance was higher in senior than junior residents, but this was not a longitudinal study of the same resident as in our study so this may have accounted for our different results. Clarke and colleagues felt this lack of improvement from second year to third year of residency may be due to a ceiling effect that obscures true differences in ability in the advanced stages of residency (Clarke et al., 2014). Perhaps our results also were due to this ceiling effect, meaning the optimal potential effect of a Simulation Wars was achieved at second year of residency so there was no improvement noted in third year of residency This is interesting because it is known that the In-Service Board scores for EM residency in the US progressively improve from first year to third year of residency. If the ceiling effect was the cause, then perhaps concentrating on Simulation Wars in residency should occur only during the first two years of residency. Another reason for the third year residents not improving on their In-Service boards from the previous year was based on their moonlighting opportunities. Our EM residency program does prohibit moonlighting if one has an In-Service score more than 10% below the average score. Based on this, a third year resident doesn't have any incentive to score well since they are graduating and will no longer need moonlighting money because their salary will drastically increase once they graduate. First and second year residents are not allowed to moonlight and since the In-Service Boards are administered midyear with their results reported two months later most second years have an incentive to study. Maybe the second year residents studied more for the In-Service than in other years. Therefore our results may not be applicable to the general public, since some programs do not even allow moonlighting.

.

Third Research Question

The third research question asked: Is there a difference in the intervention groups' In-service Board subcategory topic score according to whether they participated in a Simulation War that matched that subcategory topic? Comparing the subcategory In-Service Board scores to those who participated in a Simulation War session that matched the subcategory topic did not show a significant improvement in the majority of subcategories. There was an improvement in In-Service board scores for Thoracic Disorders if a first year resident participated in the Thoracic Simulation Wars. For second year residents who participated in the Simulation Wars on Abdominal and Gastrointestinal (GI) Simulation Wars there was an improvement in those second year residents' Abdominal and GI In-Service board scores. For third year residents who participated in the Simulation Wars on Trauma there was an improvement in the In-Service board score in the subcategory of Trauma disorders. All other types of Simulation Games did not have a significant difference in subcategory In-Service board scores. This was an important finding showing improvement in certain subcategories of In-Service Board scores, rather than no improvement in any of the subcategories. Comparing subcategory In-Service board scores by year and attendance to a particular Simulation Wars showed a significant improvement from first to third year of residency for OBGyn, and Trauma. Comparing subcategory In-Service board scores for from first year to second year there was a significant improvement for Trauma and ObGyn. Also, when comparing subcategory In-Service board scores by year and attendance for second and third years there was a significant improvement in their In-Service board score if they attended a Thoracic or Trauma Simulation Wars. In summary the subcategories of Abdominal and GI, Trauma, Thoracic, and ObGyn on In-Service Board scores were positively affected if a resident attended a Simulation Wars session in this subcategory.

We speculated that the certain subcategories where Simulation Wars participation positively affected In-Service Board scores was due to more residents attending these certain sessions. In reviewing the attendance to Thoracic, Abdominal and GI, Trauma, and ObGyn Simulation Wars where In-Service scores improved, there wasn't a consistent increased attendance for these sessions. There were more residents attending the Abdominal and GI (147) and ObGyn (111) Simulation Wars but less to the other subcategories of Thoracic (82) and Trauma (98). The most attended sessions were Abdominal and GI (147), Environmental (140) and Cardiology (114). For some reason the Thoracic and Trauma Simulation Wars, even though not more frequently attended, may have made a more memorable impact on the resident than the other subcategory Simulation War sessions. The residents were required to attend Simulation Wars unless they were on a rotation such as Medical Intensive Care or Surgical Intensive Care where they were not allowed to attend or were working the night before, not having slept for more than 12 hours. Even though residents were excused from some sessions, a large number of residents attended the Simulation Wars and were positively affected.

Clinical Implications

This is the only study that examined Simulation Wars in relation to In-Service scores and GRS scores. This was also one of two studies that examined a clinical reasoning exercise longitudinally over time in the same set of EM residents. The only other longitudinal study on clinical reasoning skill was done by Clarke and colleagues with the Ottawa Crisis Resource Management Global Rating Scale. This trend of improvement from first year to third year of residency on the ACGME evaluation competencies follows the clinical reasoning levels as one goes from novice to advanced or near expert. Clinical reasoning improves as a novice experiences more patient care and clinical situations (Swing, 2007). With the inclusion of Simulation Wars in this residency curriculum we found an improvement in clinical reasoning skills in certain subcategories of In-Service board scores. This finding has significant implications since it has not been reported in the literature. It proves that the participation in Simulation Wars improved on certain subcategory topics. Perhaps these topics are more memorable to the resident compared to the other subcategory topics, or perhaps they were the same scenario they encountered on their In-Service boards. Nevertheless, there was an improvement in certain subcategories and inclusion of these subcategory topics in Simulation Wars into other residency curriculum may also show an improvement in In-Service Board scores as a measure of clinical reasoning skills.

In an attempt to decrease medical errors, the ACGME limited the number of work hours for residents but Simulation Wars may aid the resident since there is a decrease in bedside teaching. With the restricted work hours some educators have argued that the resultant fewer hours worked in the hospital means less time to learn from patient care. Interestingly one would think the more patients seen in a fixed time frame in the emergency department, a setting where clinical reasoning is frequently practiced, one would think there would be an improvement in clinical reasoning skills. In a study by Frederick and colleagues, they sought to see if clinical productivity would correlate with initial ABEM written and oral scores for EM residents (Frederick, Hafner, Schaefer, & Aldag, 2011). Clinical productivity was defined as the number of patients seen per hours. This clinical productivity was felt to be a representative of "real –life," a measurement used for clinical performance and physician reimbursement. Their findings showed there was no significant correlation between clinical productivity and ABEM scores, with written ($r=-0.021$, $p=0.881$) or oral ($r=-0.02$, $p=0.879$) (Frederick, et al., 2011). Ledrick and colleagues also reviewed the skill of multitasking in an ED using

Relative value units (RVUs) and found that they correlated more with training level than with medical knowledge as measured by in-training scores. (Ledrick, Fisher, Thompson, & Sniadanko 2009). Relative value units (RVUs) are a measure of value used in the United States Medicare reimbursement formula for physician services. These studies probably didn't show an improvement in board scores since the resident didn't necessarily go through the process of Kolb's theory of experiential learning, where knowledge is strengthened through reflective observation, active experimentation and abstract conceptualization. In the ED there is limited time to reflect on one's action with the demand to evaluate and treat patients quickly. Through Simulation Wars one goes through this experiential learning outside the ED setting which makes it more conducive and improves clinical reaoning skills. Our finding unlike their studies did show an improvement in some subcategory topics with Simulation Wars and this may have been because this was outside the ED were experiential learning is more likely to take place.

Since any ED is where most evaluations and diagnosis are made, improving clinical reasoning skills for those who will practice there will have a huge impact. This will impact patient care, mortality/morbidity and financial cost to the patient and hospital. Extra lab testing and radiographs would not be ordered since the ED physician would have undergone clinical reasoning exercises where they could discriminate between diagnoses. The physician would most likely have a more refined differential diagnosis list which would require less testing. This would save costs incurred while in the ED. Less diagnostic errors would occur with the chance of a discharged patient returning to the ED in a deteriorated medical state, costing admittance for hospitalization or intensive care unit (ICU).

The results of our study suggest that Simulation Wars did not affect GRS scores and the majority of In-Service board scores. Despite this our study does contribute to the medical education literature. This study, unlike others that have examined the effect of a clinical reasoning exercise on In-Service board scores or GRS, we included a control group. Including a control group can measure the effect more clearly than if a control group were not present. Since In-Service board scores improve as one advances from first year to second year and then from second year to third year the effect of Simulation Wars on In-Service board scores was compared to a control. Looking at the residents; progression from year one to year two and then to year three there was an improvement on the paired t tests where an individual resident's scores were compared to their own in the intervention group.. Like the intervention group, the control group surprisingly also had an improvement from year one to year three. This finding may indicate that the clinical reasoning skills as measured on In-Service score was multifactorial so other clinical exposure, bedside teaching or didactics influenced their scores, so it was hard to determine whether it was the Simulation Wars in isolation that affected their In-Service scores. Analyzing the intervention group's In-Service subgroup category according to whether they attended a Simulation Wars in that same topic was the most revealing finding. No other study on clinical reasoning exercises did this analysis and surprisingly there was only certain topics that seemed to improve their In-Service subcategory topic. Perhaps Simulation Wars should only focus on these topic in the future. Our findings can help other residency directors tailor their clinical reasoning exercises so it will more likely improve board scores and more importantly improve clinical reasoning skills while practicing medicine on patients.

The way these Simulation Wars were formatted was in agreement with what has been recommended for clinical reasoning exercises. The Simulation Wars used in our study started with the resident being presented a scenario and then asking him/her or a small group of residents what they would do next. The student's answer was evaluated, based on whether they included the essential critical action or not. The format of the Simulation Wars was more consistent with the key-feature approach testing which is recommended where the resident may be asked after the scenario is presented "What actions will you do next?" or "What orders, if any, will you write for this patient?" It has been recommended that a clinical reasoning exercise focus on the intermediate or final outcome of the decision-making process, such as key-feature approach testing and extended-matching items (Case & Swanson, 1993; Bordage, 1987).

Implications for Future Research

The purpose of the study was to determine whether Simulation Wars affected the In-Service board scores as well as the GRS scores. There was no s significant improvement in the GRS scores under the Patient Care and Medical Knowledge competencies which were expected. There was not a significant difference when comparing the overall intervention and control groups In-Service board scores. There was a upward trend in the In service scores being higher for those attending one or more Simulation Wars for first and third year residents compared to controls in that same year of residency. This would be especially crucial for emergency medicine residencies where clinical reasoning is a major practiced skill.

Although the current study demonstrated no statistically significant difference between the GRS in the competencies of Patient Care and Medical Knowledge and In-

Service scores for the groups as a whole, further research in this area is necessary to reveal its potential effectiveness. The presentation of Simulation Wars in a different format so it is a more decision-making process such as key-feature approach testing and extended-matching items or focus on the underlying process itself such as script concordance may show more improvement in GRS and In-Service scores. Some argue that both the GRS and board scores don't adequately measure clinical reasoning but oral board scores do. Inclusion of oral board scores which are only included in the final certification boards would be helpful to truly see the effect of Simulation Wars on clinical reasoning.

Another area of future research would be to include only the ACGME recommended evaluation standards including the 1-5 Likert scale rather than other formats, such as “met expectations (ME), or below expectations (BE)”. This would help to eliminate this confounder of using different scales to evaluate residents with the GRS scales. In addition, this would allow for a better longitudinal, multicenter evaluation of educational interventions since the standard evaluation of a 1-5 Likert scale has been adopted by all residency programs in the US. Also including five scales instead of 3 scales improves precision so a significant difference in GRS scores may be shown.

Since our GRS scores and In-Service Board scores did not show a significant improvement with Simulation Wars, perhaps changing the outcome measure would be more relevant. I have talked with residents who have participated in the Simulation Wars and they feel it makes them more comfortable and confident in real life situations in their decision making and treatment so was surprised their GRS and In-Service board scores did not improve. Exploring how Simulation Wars affected their confidence and comfort level would be worthwhile. Devising a Survey that asks the resident such questions as "how valuable was Simulation wars to you?" "Did Simulation Wars help you in your decision making while working in the ED?" "Do you feel like you made quicker decisions on evaluation and treatment after participating in Simulation Wars?" This can be administered after participating in at least 2 sessions of Simulation Wars. What would even be more meaningful would the survey administered six months after graduation so we could find out if Simulation Wars was helpful in their current job as and ED physician.

Summary

The results of this study showed Simulation Wars, a clinical reasoning exercise, did not improve GRS scores in the areas expected such as Patient Care and Medical Knowledge but did show a significant improvement in certain subcategory In-Service Board scores.

According to the results of this study, the research questions were answered as follows:

1) There was not a significant difference in the Global Rating Scales scores except for the competencies of Communication and Professionalism. There was not a significant difference in the In-Service Board scores between the intervention group and control group.
2) There was a significant improvement in In-Service board scores in those who participated in Simulation Wars from Year 1 to Year 2. However, this also held

true for the control group. Thus, no improvement was evidenced after the study's intervention.

3) There was an improvement in certain subcategory topics on the In-Service board scores according to whether they were involved with a matched specific subcategory topic in the Simulation Wars.

Overall, because residents improved in their In-Service Board scores in certain subcategory topics if they participated in a matched subcategory Simulation Wars. Based on this result, it would be recommended that clinical reasoning exercises such as Simulation Wars be incorporated into emergency medicine residency curriculum.

References

Ainley, P., Rainbird, H. (2000). Apprenticeship: towards a new paradigm of learning. *34*, 324–325.

Airee, A., Guirguis, A. B., Mohammad, R. A. (2009). Clinical outcomes and pharmacists'acceptance of a community hospital's anticoagulation management service utilizingdecentralized clinical staff pharmacists. *Annals of Pharmacotherapy, 43*, 621– 628.

Aljarallah, B., Hassan, M. S. (2015). Problem based learning (PBL) vs. case based curriculum in clinical clerkship, internal medicine innovated curriculum, student prospective. *International Journal of Health Sciences*. *9*, 147–152.

American Board of Emergency Medicine. (2015). In-service scores. Retrieved from https://www.abem.org/public/docs/default-source/publication-documents/special-focus program-directors-2015.pdf?sfvrsn=6)

Andel, C., Davidow, S. L., Hollander, M, Moreno, D. A. (2012). The economics of health care quality and medical errors. *Journal of Health Care Finance*, *39*, 39–50.

Andreatta, P.1., Saxton, E., Thompson, M., Annich G. (2011). Simulation-based mock codes significantly correlate with improved pediatric patient cardiopulmonary arrest survival rates. *Pediatric Critical Care Medicine,* 12, 33-38.

Antiel, R., Thompson, S. M., Reed, D. A., James, K. M., Tilburt, J. C., Bannon, M. P., Fischer, P. R., Farley, D. R. (2010). ACGME duty-hour recommendations: A national survey of residency program directors. *New England Journal of Medicine, 363*, e12.

Antiel, R. M., Thompson, S. M., Hafferty, F. W, James, K. M., Tilburt, J. C., Fischer.P.R., Farley, D.R., Reed, D.A. (2011). Duty hour recommendations and implications for meeting the ACGME core competencies: Views of residency directors. *Mayo Clinical Proceedings*, *86*, 185–191.

Anupam, B, Schoemaker, J. L., Bhattacharya, J. (2014). Exposing physicians to reduced residency work hours did not adversely affect patient outcomes after residency. *Health Affairs*, *33*, 1832–1840.

Aspden., Wolcott, J. A., Bootman, J. L., Cronenwett, L. R. (Eds). Preventing medication errors: Quality chasm series. Washington, DC: Institute of Medicine, National Academic Press.

Atherton, J. S. (2005). Learning and teaching: Constructivism in learning. Retrieved from thttp://www.learningandteaching.info/learning/constructivism.htm.

Barsuk, J. H., Cohen, E. R., Feinglass, I., McGahie, W. C., Wayne, D. B. (2009). Use of simulation-based education ot reduce catheter-related bloodstream infection. Archives of Internal Medicien, 169, 1420-1423. .

Bartlett, K. W., Whicker, S. A., Bookman, J., Narayan, A. P., Staples, B. B., Hering, H., McGann, K, A. (2015). Milestone-based assessments are superior to Likert-type assessments in illustrating trainee progression. *Journal of Graduate Medical Education*, *7*, 75–80.

Bereiter, C, & Scardemalia, M. (1993) *Surpassing ourselves: An inquiry into the nature and implications of expertise*. Chicago, Ill: Open Court.

Bianchi, L., Gallagher, E. J., Korte, R., & Ham, H. P. (2003). Interexaminer agreement on the American Board of Emergency Medicine Oral Certification Examination. *Annals of Emergency Medicine, 41*, 859–864.

Bleakley, A. (2006). Broadening conceptions of learning in medical education: The message from teamworking. *Medical Education, 40*, 150–157.

Blum, M. G., Powers, T. W., & Sundaresan, S. (2004). Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *Annals of Thoracic Surgery, 78*, 287–291.

Bond, W., Kuhn, G., Binstadt. E., Quirk, M., Wu, T., Tews, M., Dev, P., & Ericsson, K. A. (2008). The use of simulation in the development of individual cognitive expertise in emergency medicine. Academic Emergency Medicine, 15, 1037e45.

Bordage, G. (1994). Elaborated knowledge: A key to successful diagnostic thinking. *Academic Medicine, 69*, 883–885.

Bordage, G. (1987). An alternative approach to PMPs: The 'key features' concept. Hart & R. Hardent (Eds.) Proceedings of the second Ottawa conference on further development in assessing clinical competence. (pp. 59–75). Montreal, QC: Can-Heal Publications, Inc.

Bowen, J. L. (2006). Educational strategies to promote clinical diagnostic reasoning. *New England Journal of Medicine, 355*, 2217–2225.

Bradley, P., & Postlethwaite, K. (2003). Simulation in clinical learning. *Medical Education*, *37*, 1–5.

Brauer, D., & Ferguson, K. (2015). The integrated curriculum in medical education: AMEE Guide No. 96. *Medical Teacher*, *37*, 312–322.

Brennan, T. A., Lepae, L. L., Laird N. M., Hebert, L, Localio, A. R., Lawthers, A.G., Newhouse, J. P., Weiler, P. C., & Hiatt, H. H. (1991). Incidence of adverse events and negligence in hospitalized patients. *New England Journal of Medicine, 324*, 370–376.

Bulstrode, C., Gallagher, F. A., Pilling, E. L, Furniss, D, & Proctor, R. D. (2003). A randomised controlled trial comparing two methods of teaching medical students trauma and orthopaedics: Traditional lectures versus the "donut round." *Surgeon*, *1*, 76–80.

Carroere, B., Gagnon, R., Charlin, B., Downing, S., & Bordage, G. (2009). Assessing clinical reasoning in pediatric emergency medicine: Validity evidence for a script concordance test. Annals of Emergency Medicine, 53, 647–652.

Case, S. M., & Swanson, D. B. (April 1989). Evaluating diagnostic pattern: A psychometric comparison of items with 15, 5, and 2 options. Presented at the meeting of the American Educational Research Association, San Francisco, CA.

Case, S. M., & Swanson, D. B. (1993). Extending matching items: A practical alternative to free response questions. *Teaching and Learning in Medicine*, *5*, 107–115.

Chapman, D. M., Car, D. M., & Aubin, C. D. (2005). Clinical decision making. In J. A. Mars (Ed.), *Rosen's Emergency Medicine: Concepts and Clinical Practice* (6th ed.), (pp. 125–133), St. Louis, MO: Mosby.

Charlin, B., Tardif, J., & Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine*, *75*, 182–190.

Charlin, B., Brailovsky, C., Leduc, C., & Blouin, D. (1998). The diagnostic script questionnaire: A new tool to assess a specific dimension of clinical competence. *Advances in Health Sciences Education*, *3*, 51–58.

Charlin, B., Tardif, J., & Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine*, *75*, 182–190.

Clarke, S., Horeczdo, T., Carlisle, M., Barton, J. D., Ng, V., Al-Somali, T., & Bair, A. E. (2014). Emergency medicine resident crisis resource management ability: A simulation-based longitudinal study. *Medical Education Online*, *19*(1), 1–11.

Cohen-Gadol, A. A., Piepgras, D. G., Krishnamurthy, S., & Fessler, R. D. (2005). Resident duty hours reform: Results of a national survey of the program directors and residents in neurosurgery training programs. *Neurosurgery*, *56*, 398–403.

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hamstra, S. T. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *Journal of American Medical Association*, *306*, 978–988.

Cook, R. L., & Woods, D. O. (1994). Opening at the sharp end: The complexity of human error. In M. S. Bogner (Ed.), *Human Error in Medicine* (pp. 255–309). Mahwah, NJ: Lawrence Erlbaum.

Costa, M. L, van Rensburg, L., & Rushton, N. (2007). Does teaching style matter? A randomized trial of group discussion versus lectures in orthopaedic undergraduate teaching. *Medical Education*, *41*, 214–7.

Council of Emergency Medicine Residency Directors (CORD). Retrieved from http://www.cordem.org/i4a/pages/index.cfm?pageid=3403 2015.

Crofts, J. F., Fox, R., Ellis, D., Winter, C., Hinshaw, K., & Draycott, T. J. (2008). Observations from 450 shoulder dystocia simulations: Lessons for skills training. *Obstetrics and Gynecology, 112*, 906–912.

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine*, *84*, 1022–1028.

Croskerry, P. (2005). The theory and practice of clinical decision making. *Canadian Journal of Anesthesia, 52,* R1–8.

Daniel, K., Lipman, S., Harney, K., Arafeh, J., & Druzin, M. (2008). Use of simulation based team training for obstetric crises in resident education. Simulation Health, 3, 154–160.

Drake, R. L. (2014). A retrospective and prospective look at medical education in the United States: Trends shaping anatomical sciences education. *Journal of. Anatomy*, *224*, 256–260.

Draycott, T. J., Crofts, I. F., Ash, I. P, Wilson, L.V., Yard, E., Sibanda, T., & Whitelow, A. (2008). Improving neonatal outcome through practical shoulder dystocia training. *Obstetrics and Gynecology,112*, 14–20.

Draycott, T., Sibanda,T., Owen, I., Akanda, V., Winter, C., Reading, S., & Whitelow, A. (2006). Does training in obstetric emergencies improve neonatal outcome? *British Journal of Obstetrics and Gynecology, 113*, 177–182.

Dunn, M. M., & Woolliscroft, J. O. (1995). Assessment of surgical pattern recognition examination. *American Journal of Surgery,* 169, 341–343.

Durning, S. J., Dong, T., Artino, A. R., der Vleuten, C. V., Holmboe, E., & Schuwirth, L. (2004). Dual processing theory and experts' reasoning: Exploring thinking on national multiple-choice questions. *Perspective Medical Education*, *4*, 168–175.

Eddy, D. M. (1996). *Clinical decision making. From theory to practice: A collection of essays published by the American Medical Association.* Wood Hole, MA: Jones and Bartlett Publishers.

Eiser, A. R., & Connaughton-Storey, J. (2008). Experiential learning of systems-based practice: A hands- on experience for first-year medical residents. *Academic Medicine*, *83*, 916–923.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction.* American Philosophical Association. ERIC document. Millbrae, CA: The California Academic Press. (ERIC No. ED 31542)

Fairbanks, R. J., Bisantz, A. M., & Sunm, M. (2007). Emergency department communication links and patterns. *Annals of Emergency Medicine, 50*, 396–406.

Farheinkopt, A. M., Sectish, T. C., Berger, L. K., Sharek, P. J., Lewin, D, Chiang, V. W., Edwards, S., & Landrigan, C. P. (2008). Rates of medication errors among depressed and burnt out residents: Prospective cohort study. *British Medical Journal, 336*, 488–491.

Feinstein, A. R. (1967). Clinical Judgment. Science, 13, 246-247

Fernandez, G. L., Page, D. W., Coe, N. P., Lee, P. C., Patterson, L. A., Akylizard, L., St. Louis, M., Amaral, M. H, Wait, R. B, & Seymour, N. E. (2012). Boot camp: Educational outcomes after 4 successive years of preparatory simulation-based training at onset of internship. *Journal of Surgical Education*, *69*, 242–248.

Frederick, C., Hafner, D., Schaefer, T. J., Aldag, C. (2011). Outcome measures of emergency medicine residency graduates: Do measures of academic and clinical performance during residency training correlate with American Board of Emergency Medicine test performance? *Academic Emergency Medicine*, *18*, S59–S64.

Freed, G. L., Dunham, K. M., & Lamarand, K. E. (2009). Research Advisory Committee of the American Board of Pediatrics. Hospitalists'involvement in pediatrics training: Perspectives from pediatric residency program and clerkship directors. *Academic Medicine*, *84,* 1617–1621.

Graber, M., Gordon, R., & Franklin, N. (2002). Reducing diagnostic errors in medicine: What's the goal? *Academic Medicine, 77*, 981–92.

Graber, M. L., Franklin, N., & Gordon, R. (2005) .Diagnostic error in internal medicine. *Archives of Internal Medicine*, *165*, 1493–9.

Gray, J. D. (1996). Global rating scales in residency education. *Academic Medicine*, *71*, S55–S63.

Grześkowiak, M. (2006). The effects of teaching basic cardiopulmonary resuscitation—A comparison between first and sixth year medical students. *Resuscitation*, *68*, 391–397.

Guile, D., & Young, M. (2001). Apprenticeship as a conceptual basis for a social theory of learning. In C. Paechter, M. Preedy, D. Scott, & J. Soler (Eds.) *Knowledge, Power and Learning* (pp. 56-73). London: Paul Chapman.

Hamlin, S., Cutler,T., Hafner, F. R., & Thompson, L. (2008). Correlation of emergency medicine residency training simulated oral board examinations scores with national ABEM oral examination scores. *Academic Emergency Medicine*, *15*, s54.

Haurani, M. J., Rubinfeld, I., Rao, S., Beaubien, J., Musial, J. L., Parker, A., Reicker, C., Raafat A., & Shepard, A. (2007). Are the communication and professionalism competencies the new critical values in a resident's global evaluation process? *Journal of Surgical Education*, *64*, 351–356.

Heckmann, J. G., Bleh, C., Dutsch, M., Lang, C. J., & Neundorfer, B. (2003). Does improved problem-based teaching influence students' knowledge at the end of their neurology elective? An observational study of 40 students. *Journal of Neurology*, *250*, 1464–8.

Hedrick, T. L., & Young, J. S. (2008). The use of "war games" to enhance high-risk clinical decision-making in students and residents. *American Journal of Surgery*, 195, 843–849.

Heffer, R. E., & Slater, C. H. (1971). Measuring the process of solving clinical diagnostic problems. *British Journal of Medical Education*, *5*, 48–52.

Holmboe, E. S, Lipner, R. S, & Greiner, A. (2000). Assessing quality of care, knowledge matters. *Journal of American Medical Association*, 299, 338–340.

Humbert, A. I., Besinger, G., & Miech, E. J. (2011). Assessing clinical reasoning skills in scenarios of uncertainty: Convergent validity for a Script Concordance Test in an emergency medicine clerkship and residency. *Academic Emergency Medicine. 18*, 627–34.

Hung, D. (2002). Situated cognition and problem-based learning: Implications for learning and instruction with technology. *Journal of Interactive Learning Research, 13,* 393–415.

Immerman, I., Kubiak, E. N., & Zuckerman, J. D. (2007). Resident work-hour rules: A survey of residents' and program directors' opinions and attitudes. *American Journal of Orthopedics, 36,* e172–179.

Institute of Medicine. (2000). *To err is human: Building a safer health system.* Washington DC: National Academies Press.

Institute of Medicine. (2015). *Improving diagnosis in healthcare*. Retrieved from http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2015/Improving- Diagnosis/DiagnosticError_ReportBrief.pdf

Issenberg, S. B., McGaghie, W. C., Hart, J. R., Mayer, J. W., Felner, J. M., Petrusa, E. R., Waugh, R, A., Brown, D. D., Safford, R. R., Gessner, I. H., Gordon, D. L., & Ewy, G. A. (1999). Simulation technology for health care professional skills training and assessment. *Journal of American Medical Association, 282,* 861–866.

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee, G. D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Medical Teacher, 27,* 10–28.

James, J. T. (2013) A new, evidence-based estimate of patient harms associated with hospital care. *Journal of Patient Safety 9*, 122–128.

Johannsson, H., Ayida, G., & Sadler, C. (2005). Faking it? Simulation in the training of obstetricians and gynaecologists. *Current Opinion in Obstetrics and Gynecology, 17,* 557.

.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds).(1999). *To err is human: Building a safer health care system. Report of the Institute of Medicine*. Washington, DC: National AcademyPress.

Kolb, D, A. (1976). *Learning style inventory. Technical Manual*. Boston, MA: McBer.

Kolb, D. A., & Fry, R. (1975). Toward an applied theory of experiential learning. In C. Cooper (Ed.), *Theories of Group Process.* London: John Wiley.

Kolb, D. A.(1983). *Experiential learning: Experience as the source of learning and development*. Upper Saddle River, NJ: Prentice-Hall.

Krishnamurthy, S., Satish, U., Foster, T., Streufert, S., Dewan, M., & Krummel, T. (2009). Components of critical decision making and absite assessment: Toward a more compehensive evaluation. *Journal of Graduate Medical Education*, *1*, 273–7.

Kupferman, T. A., & Lian, T. S. (2005). Implementation of duty hour standards in otolaryngology head and neck surgery residency training. *Otolaryngology – Head and Neck Surgery*, *132*, 819–822.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation* (pp. 27–44), Cambridge, UK: Cambridge University Press.

Leape, L. L., Brennan, T. A., Laird, N., Lawthers, A. G., Localio, A.,R., Barnes, B. A., Hebert, L., Newhouse, J. P., Weiler, P. C., & Hiatt, H. (1991). The nature of adverse events in hospitalized patients: Results of the Harvard Medical Practice Study II. *New England Journal of M*edicine, *324*, 377–384.

Ledrick, D., Fisher, S., Thompson, J., & Sniadanko, M. (2009). An assessment of emergency medicine residents' ability to perform in a multitasking environment.

Academic Medicine, 84, 1289–1294.

Ledue, Chelsey. (2010, August 9). Study: Medical errors cost U.S. economy almost $20 billion in '08. *Health Care Finance News.* Retrieved from http://www.healthcarefinancenews.com/news/study-medicalerrors-cost-us-economy almost10billion-08.

Liebermanv, J. D., Olenwine, J. A., Finley, W., & Nicholas, G. G. (2005). Residency reform: Anticipated effects of ACGME guidelines on general surgery and internal medicine residency programs. *Current Surgery, 62*, 231–236.

Littlefield, J., Paukert, J., & Schoolfield, J. (2001). Quality assurance data for residents' global performance ratings. *Academic Medicine, 76,* S102–S104.

Lloyd, J. (2013,March 25). Studies: Residents make more errors on shorter shifts. USA TODAY. Retrieved from http://www.usatoday.com/story/news/health/2013/03/25/residents-health-medical-errors-hospitals/2018873/

Louw, G., Eizenberg, N., & Carmichael, S. W (2009). The place of anatomy in medical education: AMEE guide no.41. *Medical Teacher, 31*, 373–386.

Macedonia, C .R., Gherman, R. B., & Satin, A. J. (2003). Simulation laboratories for training in obstetrics and gynecology. *Obstetrics and Gynecology, 102*, 388–392.

Makary, M. A. (2016) Medial error-the third leading cause of death in the US. *British Medical Journal*, *353*:12139.

Marcum, J. A. (2012). An integrated model of clinical reasoning: Dual-process theory of cognition and metacognition. *Journal of Evaluation and Clinical Practice*, *18*, 954–961.

Maslovitz, S., Barkai, G., Lessing, J. B., Ziv, A., & Many, A. l. (2007). Recurrent obstetric management mistakes identified by simulation. *Obstetrics and Gynecology, 109*, 1295–1300.

Meterissain, S., Zabolotny, B., Gagnon, R., & Charlin, B. (2007). Is the script concordance test a valid instrument for assessment of intraoperative decision-making skills? *American Journal of Surgery*, *193*, 248–251.

Mickelsen, S., McNeil, R., Parikh, P., & Persoff, J. (2011). Reduced resident “code blue” experience in the era of quality improvement: New challenges in physician training. *Academic Medicine, 86*, 726–730.

Miettinen, R. (2000). Andragogy in action. Applying modern principles of adult education. *International Journal of Lifelong Education, 19*, 54–72.

Morgan, P. J. Cleave-Hogg, D. (2005). Simulation technology in training students, residents and faculty. *Current Opinion Anaesthesiology*, 18, 199-203.

Munger, B. S., Krome, R. L., Maatsch. J. C., & Podgomy, G. (1982). The certification examination in emergency medicine: An update. *Annals of Emergency Medicine*, *11*, 91–96.

Nasca, T. J., Day, S. H., & Amis, E. S., Jr. (2010). ACGME Duty Hour Task Force. The new recommendations on duty hours from the ACGME Task Force. *New England Journal of Medicine, 363*, e3.

Norman, G. R., Young, M., & Brooks, L. R. (2007). Non-analytical models of clinical reasoning: The role experience. *Medical Education, 41*, 1140–1145.

Norman, G. (2009). Dual processing and diagnostic errors. *Advances in health. sciences education: Theory and practice*. *14*, 137–49.

Nuthalapaty, F. S, Carver, A. R., Nuthalapaty, E. S., & Ramsey, P. S. (2006). The perceived impact of duty hour restrictions on the residency environment: A survey of residency program directors. *American Journal of Obstetrics and Gynecology, 19*, 1556–1562.

O'Malley, C. W., Emanuele, M., Halasyamani, L., & Amin, A.N. (2008). Society of Hospital Medicine Glycemic Control Task Force. Bridge over troubled waters: Safe and effective transitions of the inpatient with hyperglycemia. *Journal of Hospital Medicine*, *3*, 55– 65.

Pabst, R., Westermann, J, & Lippert., H. (1986). Integration of clinical problems in teaching gross anatomy: Living anatomy, X-ray anatomy, patient presentation on films depicting clinical problems. *The Anatomical Record*, *215*, 92-94.

Papp, K. K., Huang, G. C., Lauzon, L. M, Delva, D., Fischer, M., Konopasek, L., Schwartzstein, R. M., & Gusic, M. (2014). Milestones of critical thinking: A developmental model for medicine and nursing. *Academic Medicine*, *89*, 715–720.

Papanna, K. M., Kulkarni, V., Tanvi, D., Lakshmi, V., Kriti, L., Unnikrishnan, B., Akash, S., Tejesh, S., & Sumit Kumar, S. (2013). Perceptions and preferences of medical students regarding teaching methods in a medical college, Mangalore India. *African Health Sciences*, *13*(3), 808–13.

Patel, V. L, Arocha, J. F, & Kaufman, D. R. (1994). Diagnostic reasoning and medical expertise. *Psychology of Learning and Motivation: Advances in Research and Theory*, *31*, 237–252.

Patel, V. L., Groen, G. J., & Patel, Y. C. (1997). Cognitive aspects of clinical performance during patients workup: The role of medical experts. *Advanced Health Science Education, 2,* 95–114.

Pelaccia, T. , Tardif , J., Triby, E., & Charlin, B. (2011). An analysis of clinical reasoning through a recent and comprehensive approach: The dual-process theory. Medical Education Online ,16.

Perkins, G. D. (2007). Simulation in resuscitation training. *Resuscitation, 7,* 202–211.

Philibert, I., Friedman, P., & Williams,W. T. (2002). New requirements for resident duty hours. *Journal of the American Medical Association*, *288*, 1112–1114.

Pronovost, P. J., Thompson, D. A., & Holzmueller, C. G. (2005). Defining and measuring patient safety. *Critical Care Clinics*, *21*, 1–18.

Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*, 1–12.

Reid, R. (2015). The history of chess. Retrieved from http://www.chess-poster.com/English/chesmayne/history_of_chess.htm

Resnick, A. S., Todd, B. A., Mullen, J. L., & Morris, J. C. (2006). How do surgical residents and non-physician practitioners play together in the sandbox? *Current Surgery,63*, 155–164.

Robinson, M., Mogensen, K. M., Grudinskas, G. F., Kohler, S., & Jacobs, D. O. (2005). Improved care and reduced costs for patients requiring peripherally inserted central catheters: The role of bedside ultrasound and a dedicated team. *Journal of Parenteral and Enteral Nutrition*, *25*, 374–379.

Sandhu, H., Carpenter, C., Freeman, K., Nabors, S .G., & Oson, A. (2005) Clinical decision making: Opening the black box of cognitive reasoning. *Annals of Emergency Medicine, 48*, 713–719.

Savoldelli, G. I., Naik, V. N., Joo, H. S., Houston, D. L., Graham, M., Yee, B., & Hamstra, S. T. (2006). Evaluation of patient simulator performance as an adjunct to the oral examination for senior anesthesia resident. *Anesthesiology, 104*, 475–481.

Schmidt, H. G., & Boshuizen, H. P. A. (1992). On acquiring experts in medicine. *Educational Psychology Review*, *5*, 205–221.

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education*, *41*, 1133–1139.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1992). A cognitive perspective on medical experts: Theory and implications. *Academic Medicine, 65*, 611–621.

Schumaker, D. (2014). Domain of competence: Patient care. *Academic Pediatrics, 14*:S13-S35.

Sen, S., Krangler, H. R., Didwania C. K., Schwartz, A .C, Amarnath, S., Kolars, J. C., Dalach, G. W., Nichols, B., & Guille, C. (2013). Effects of the 2011 duty hour reforms on interns andtheir patients. *Journal of American Medical Association: Internal Medicine*, *173*, 657–662.

Seymour, N. E., Gallagher, A. G., Roman, S. A., O’Brien, M. K., Bansal, V. K., Andersen D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery, 236*, 458–463.

Shojania, K. G., Burton, E. C., McDonald, K. M., & Goldman, L. (2003). Changes in rates of autopsy-detected diagnostic errors over time: A systematic review. *Journal of American Association of Medicine*, *289*, 2849–2856.

Smith, C. M., Perkins, G. D., Bullock, I.,& Bion, J. F. (2007). Undergraduate training in the care of the acutely ill patient: A systematic literature review. *Intensive Care Medicine, 33*, 901– 907.

Society of Acturies. (2010). Society of Actuaries study finds medical errors annually cost at least $19.5 billion nationwide. Retrieved from https://www.soa.org/News-and-Publications/Newsroom/Press-Releases/Society-of-Actuaries-Study-Finds-Medical-Errors-Annually-Cost-at-Least-$19-5-Billion-Nationwide.aspx

Sos, J. C., Jr., Blatt, M. A., Higgins, M. D., & Morton, K. J. (1998). *Medical decision-making.* Boston, MA: Butterworths.

Stiegler, M. P., Neelankavi, C., Canales, C., & Dhillon, A, (2011). Cognitive errors detected in anaesthesiology: A literature review and pilot study. *British Journal of Anaesthesia, 108*, 229–235.

Stiegler, M. P., Neelankavil, J. P., Canales, C., Dhillon, A. (2012). Cognitive errors detected in anaesthesiology: a literature review and pilot study. British Journal of Anaesthesiology. 108, 229-235.

Sugand, K., Abrahams, P., & Khurana, A. (2010). The anatomy of anatomy. A review for its modernization. *Anatomical Science Education, 3*, 83–93.

Swanson, D. B., Norcini, J. J., & Grosso, L. J. (1987). Assessment of clinical competence: Written and computer-based simulations. *Assessment and Evaluation in Higher Education*, *12*, 220–46.

Swing, S. R. (2007). The ACGME outcome project: Retrospective and prospective. *MedicalTeacher*, *29*, 648–54.

Swing, S. R., Clyman, S. G., Holmboe, E. S, & Williams, R. G. (2009). Advancing resident assessment in graduate medical education. *Journal of Graduate Medical Education*, *1*, 278–86.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychol*ogy, *4*, 25–29.

Uwe, R. (December 10, 2010). The little-known decision-makers for medicare physician fees. *The New York Times*. Retrieved July 6, 2011.

Ulmer, C., Wolman, D., & Johns, M. (Eds.). (2008). *Resident duty hours: Enhancing sleep, supervision and safety*. Washington, DC: National Academies Press.

Vygotsky, L. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wayne, D. B., Didwania, A., Feinglass, J., Fudala, M. J., Bursuk, J. H., & McGaghie, W. C. (2008). Simulation-based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. *Chest*, *133*, 56–61.

Weinstein, M. C. Fineberg, H. V., & Elstein, A. S. (1980). *Clinical decision analysis.*

Philadelphia, PA: WB Saunders.

West, C. P., Tan, A. D., Habermann, T. M., Sloan, J. A., & Shanafelt, T. D. (2009). Association of resident fatigue and distress with perceived medical errors. *Journal of American Medical Association, 302*, 1294–300.

West, C. P., Huschka, M. M., Novotny, P. J., Sloan, J. A., Kolars, J. C., Habermann, T. M., & Shanafelt,T. D. (2006). Association of perceived medical errors with resident distress and empathy, a prospective longitudinal study. *Journal of American Medical Association, 296*, 1071–1078.

White, A. A., Gallagher, T. H., Krauss, M. J., Garbutt, J., Waterman, A. D., Dunagan, W. C., Fraser, V. J., Levinson, W., & Larson, E. B. (2008). The Attitudes and experiences of trainees regarding disclosing medical errors to patients. Academic Medicine. 83, 250–256.

Xu, T., Xu, J., Yu, X., Ma, & Wang, Z. (2012). Clinical decision-making by the emergency department resident physicians for critically ill patients. *Frontier Medicine, 6*, 89–93.

Young, J. S., Stokes, J. B., Denlinger, C. E., & Dubose, J. E. (2007). Proactive versus reactive: The effect of experience on performance in a critical care simulator. *American Journal of Surgery, 193*, 100–104.

Young, J. S., DuBose, J. E., Hedrick, T. L., Conaway, M. R., & Nolley, B. (2007) The use of “war games” to evaluate performance of students and residents in basic clinical scenarios: A disturbing analysis. *Journal of Trauma, 63*(3), 556–564.

Ziv, A., Wolpe, P. R., Small, S. D., & Glick, S. (2003). Simulation-based medical education: An ethical imperative. *Academic Medicine, 78*, 768–788.

Appendix A

Example of Simulation Wars Scenario

(Appendix A document)
Case: “My heart is racing”
Case designed for: All residents
Difficulty: Moderate difficulty

Primary Objectives:
1. Demonstrate the appropriate workup for a patient with a wide-complex tachycardia
2. Recognize an irregular wide-complex tachycardia on EKG
3. Differentiate between the different causes of a wide-complex tachycardia
4. Recognize the EKG features of a patient with Wolff Parkinson White Syndrome
5. Appropriately treat a patient with wide-complex tachycardias

Critical Actions
1. Places patient on pacer pads
2. Orders an EKG
3. Ask for EKG from previous visit
4. Recognize an irregular wide-complex tachycardia on EKG
5. Use procainamide or amiodarone to convert rhythm
6. Avoid usage of AV nodal blocking agents for rate control

EXAMINER ONLY

History
HPI:
28 year old previously healthy male presenting to the ED for episodes of “heart racing.” He has been having several instances over of his heart racing over the past few days. Symptoms usually last one hour and usually stop on their own. This current episode began approximately three hours prior to arrival, without resolution. Patient is unable to determine in what context the symptoms begin. Episodes have occurred with both exercising and while at rest.
PMHx:
None
(if asked about previous episodes, he will report a visit to the ED several years ago after an episode resolved. He was told to “see his doctor,” which he never did)
Medications: None
Allergies: NKDA
Family Hx: None
Social Hx:
History of occasional smoking and ETOH use. Recreational marijuana use.
ROS:
+lightheadedness, +occasional SOB, +nausea
- respiratory distress, -chest pain, -diaphoresis, -abdominal pain, -vomiting, -syncope, -recent illness, -extremity pain or swelling

EXAMINER ONLY
Physical Exam:

VS: HR 120s -140s (irregular, wide complex), BP 115/92, RR 18, O2 98% RA
General: Healthy appearing. Awake, alert, oriented. Comfortable without distress
HEENT: No evidence of trauma, NCAT. Pupils equal, round, reactive. Extra-ocular movements intact. Ears normal. Tympanic membranes normal. Mucous membranes moist.
Cardiovascular: Tachycardia, with irregular rate and rhythm. Equal pulses in all extremities. No murmurs, rubs, gallops.
Lungs: Clear to auscultation. No wheezes, crackles, rales. Equal breath sounds.
Abdomen: Soft, nontender and nondistended. No organomegaly. Normal bowel sounds.
Extremities: No muscle tenderness. No edema or swelling. Full range of motion.
Skin: Pale in appearance. No diaphoresis. No rashes, petechia, or purpura.
Neurological: Normal mental status. Cranial nerves intact. No sensory or motor deficits. Normal gait.

EXAMINER ONLY
Instructor Notes
1. Requests for old records: Only if requested, the patient's old EKG from a prior visit will be present and provided. It will have a short PR interval, and delta wave diagnostic for WPW
2. Change in patient's conditions: Patient will remain stable unless any of the following occur
a. IV fluid administration - BP will increase slightly, but no change in other vital signs
b. Vagal maneuvers - Carotid massage and valsalva will have no effect on patient's condition or vitals
c. Magnesium sulfate - No effect
d. AV nodal blocker (CCB, BB, Adenosine, Digoxin) will lead to patient deterioration. Heart rate will increase and then will degenerate into ventricular fibrillation (patient will lose pulses and patient will become unresponsive). Will need to begin ACLS.
e. Synchronized cardioversion - Will convert the patient into sinus rhythm at a rate of 80-90 bpm.
f. Defibrillation - Will result in deterioration into ventricular fibrillation with loss of pulses and the patient becoming unresponsive. Participants will need to begin ACLS.
g. Procainamide or Amiodarone - Procainamide at 20-30mg/min with a max dose of 17mg/kg or Amiodarone 150mg over 10 min will terminate the wide-complex tachycardia and result in a sinus rhythm at a rate of 80-90 beats per minute.

WPW PEARLS
- 0.2% of patients
- up to 20% of patients with congenital defects
- Sudden cardiac death 4%
- Bundle of Kent - accessory pathway that bypasses the AV node
- no decremental conduction

- impulses reach ventricle before those conducted through the AV node
- palpitations most common symptom (dizziness, syncope, sob)
- Reentry circuit
- 90% orthodromic (narrow complex) - anterograde through normal pthwy and retrograde conduction through accessory pthwy
- 10% antidromic (wide complex) - anterograde via accessory pathway

Appendix B

IRB Approval Letters

(Appendix B Document)

UTHealth
The University of Texas
Health Science Center at Houston

Committee for the Protection of Human Subjects
6410 Fannin Street, Suite 1100
Houston, Texas 77030

NOTICE OF APPROVAL TO IMPLEMENT REQUESTED CHANGES

March 02, 2016

HSC-MS-15-0811 - The Effect of Using Simulation Games on Emergency Medicine Residents' Reasoning Skills
PI: Dr. Donna Mendez

Reference Number: 134182

PROVISIONS; Unless otherwise noted, this approval relates to the research to be conducted under the above referenced title and/or to any associated materials considered at this meeting, e.g. study documents, informed consent, etc.

APPROVED: By Expedited Review and Approval

CHANGE APPROVED: Identifiers will be retained through the matching process and then destroyed

REVIEW DATE: March 2, 2016

APPROVAL DATE: March 2, 2016

CHAIRPERSON: L. Maximilian Buja, MD L. Maximilian Buja

Upon receipt of this letter, and subject to any provisions noted above, you may now implement the changes approved.

CHANGES: The principal investigator (PI) must receive approval from the CPHS before initiating any changes, including those required by the sponsor, which would affect human subjects, e.g. changes in methods or procedures, numbers or kinds of human subjects, or revisions to the informed consent document or procedures. The addition of co-investigators must also receive approval from the CPHS. ALL PROTOCOL REVISIONS MUST BE SUBMITTED TO THE SPONSOR OF THE RESEARCH.

INFORMED CONSENT: Informed consent must be obtained by the PI or designee(s), using the format and procedures approved by the CPHS. The PI is responsible to instruct the designee in the methods approved by the CPHS for the consent process. The individual obtaining informed consent must also sign the consent document. **Please note that if revisions to the informed consent form were made and approved, then old blank copies of the ICF MUST be destroyed. Only copies of the appropriately dated, stamped approved informed consent form can be used when obtaining consent.**

UNIVERSITY of HOUSTON

DIVISION OF RESEARCH

March 4, 2016

Dr. Donna Mendez
c/o Dr. Sara G. McNeil
Curriculum and Instruction

Dear Dr. Donna Mendez,

Based upon your request for exempt status, an administrative review of your research proposal entitled "The Effect of Using Simulation Games on Emergency Medicine Residents' Reasoning Skills" was conducted on December 16, 2015.

At that time, your request for exemption under **Category 4** was approved pending modification of your proposed procedures/documents.

The changes you have made adequately respond to the identified contingencies. As long as you continue using procedures described in this project, you do not have to reapply for review. * Any modification of this approved protocol will require review and further approval. Please contact me to ascertain the appropriate mechanism.

If you have any questions, please contact Alicia Vargas at (713) 743-9215.

Sincerely yours,

Kirstin Rochford, MPH, CIP, CPIA
Director, Research Compliance

*Approvals for exempt protocols will be valid for 5 years beyond the approval date. Approval for this project will expire **March 1, 2021**. If the project is completed prior to this date, a final report should be filed to close the protocol. If the project will continue after this date, you will need to reapply for approval if you wish to avoid an interruption of your data collection.

Protocol Number: 16197-EX