Integration of Simulation into Healthcare Education through Applied Constructivism: A Randomized, Switching Replications Experiment

Samuel A. Yoders
Nova Southeastern University, flyoders2@yahoo.com

This document is a product of extensive research conducted at the Nova Southeastern University College of Engineering and Computing. For more information on research and degree programs at the NSU College of Engineering and Computing, please click here.

Follow this and additional works at: https://nsuworks.nova.edu/gscis_etd
Part of the Computer Sciences Commons

NSUWorks Citation
https://nsuworks.nova.edu/gscis_etd/1002.

This Dissertation is brought to you by the College of Engineering and Computing at NSUWorks. It has been accepted for inclusion in CEC Theses and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
Integration of Simulation into Healthcare Education through Applied Constructivism: A Randomized, Switching Replications Experiment

by

Samuel A. Yoders

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computing Technology in Education

College of Engineering and Computing
Nova Southeastern University
2017
Approval Page

We hereby certify that this dissertation, submitted by Samuel Yoders, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

_____________________________________________  __________________
Gertrude W. Abramson, Ed.D.
Chairperson of Dissertation Committee

Date

_____________________________________________  __________________
Martha M. Snyder, Ph.D.
Dissertation Committee Member

Date

_____________________________________________  __________________
Marilyn Olander, Ph.D.
Dissertation Committee Member

Date

Approved:

_____________________________________________  __________________
Yong X. Tao, Ph.D., P.E., FASME
Dean, College of Engineering and Computing

Date

College of Engineering and Computing
Nova Southeastern University

2017
This report describes the development, deployment, and analysis of an experimental instructional unit using applied constructivism instructional design (ID). The ID template was used to integrate a high-fidelity simulator into an undergraduate healthcare degree curriculum in a private, not-for-profit university. A switching-replications experimental design was used with random assignment of volunteer participants to initial treatment and control groups. Quantitative analysis of learning outcomes using standardized assessments was performed, including correlational analysis for knowledge transfer of simulator skills to clinical skills.

Statistically significant positive effects were found for the educational outcomes of participants when measuring both the knowledge and application of heart anatomical structures and views for examination of the heart with ultrasound. Mild positive correlations were found between performance on the simulator and performance in an actual clinical setting, with limited predictive value between the two. The switching-replications experimental design helped to control for potentially strong social effects that could have endangered internal validity and to maximize the data available for analysis.

Many of the constructivist-based ID features of the educational unit resulted in positive feedback and participation from participants. However, cautionary findings relating to the ID features included the need to carefully evaluate their use, as there was a tendency for participants to not value the performance of certain features if they were not going to be graded, despite their likely educational benefit.

Future research suggested includes repetition across similar institutions with disparate student populations, and use of the educational unit ID template to implement simulation technology in other educational realms. Other possibilities include determining the effects on learning outcomes of a more-realistic user interface (UI) design and/or increased realism (difficulty) in the simulation itself. Related qualitative-based research could include structured interviews to determine participant satisfaction and learning Outlooks, and investigation of the learners’ thoughts and perceptions as they use actual ultrasound machines after practicing on the simulator through think-aloud and active interview techniques.
Acknowledgements

I would like to acknowledge and thank Dr. Gertrude (Trudy) Abramson, Professor Emerita, for her guidance, encouragement, and excellent advice throughout the dissertation process. Dr. Abramson, you are truly an inspiration, and it was an honor and a privilege to have worked with you. Heartfelt thanks to Dr. Marilyn Olander and Dr. Martha Snyder for their expert guidance and efforts on the dissertation committee. Thank you both so very much for your time and support, and especially for your kind words of encouragement throughout. Thank you to Dr. Steven Terrell for re-awakening my appreciation of quantitative statistical analysis, and for his very helpful book on that topic which served as an important resource in this effort.

Thank you to Dr. Stan Timofeev and Professor Bruce Martin for their tremendous help as subject-matter experts in echocardiography. Providing encouragement and ongoing support were my Department Chair, Dr. Brianna Kent, my Assistant Dean, Dr. Sandrine Gaillard-Kenney, and the Dean of the College of Health Care Sciences, Dr. Stanley Wilson. Thank you all for making it possible for me to pursue this opportunity.

Special thanks to all of my family and friends for all of their words of wisdom and unwavering optimism, especially my parents Harry and Sara Yoders, my children Hope and Harry Yoders, and my good friend Zoltan Zlodos. The biggest thanks of all however, go to Holly, my wife of 26 years, for her loving support and encouragement throughout this five-year journey. I could not have done this without her by my side.
# Table of Contents

Approval Page   ii  
Abstract iii  
Acknowledgements iv  
List of Figures vii  

Chapters  
1. **Introduction** 1  
   Background 1  
   Problem Statement 4  
   Dissertation Goal and Hypothesis 5  
   Research Questions 7  
   Limitations 8  
   Definitions and Acronyms 10  

2. **Review of the Literature** 16  
   Simulation Use in Undergraduate and Healthcare Education 17  
   Integration of Simulation into Healthcare Education Curriculum 20  
   Instructional Design in Healthcare Education Using Simulation 22  
   Constructivist Design in Education 26  
   Constructivism Discussion 28  
   Key Theorists of Interest 30  
   Two Foundational Constructivist Concepts 31  
   Application in the 21st Century 32  
   Summary 33  

3. **Methodology** 35  
   Research Design 36  
   The Experimental Educational Unit 42  
   ID Conception: A Comprehensive Lesson Plan 47  
   ID Implementation: A Focused Lesson Plan, Later Revised 48  
   Assessment Within the Instructional Unit 49  
   Addressing the Research Questions 52  
   Instruments 53  
   Data Collection and Analysis 54  
   Resources 56  
   Execution 57  
   Summary 60  

4. **Results** 67  
   Limited Results of the Pilot Run-through 68  
   Full Experiment: Baseline Assessment Analysis Between Groups 72  
   Between Group Assessment Analysis Round One and Two 74  
   Summary of Assessment Results Between Groups 78
Within-group Assessment Analysis Results Round One  80
Within-group Assessment Analysis Results Round Two  82
Summary of Assessment Results Within Groups  83
Additional Confirmation of Significance for Round One  85
Correlational Analysis  86
Summary of Correlational Analysis  91

5. Conclusions, Implications, Recommendations, and Summary  92
   Conclusions  92
   Implications  101
   Recommendations  106
   Summary  108

Appendices
A. Comprehensive four week long Lesson Plan.  126
B. Focused, three-week long Lesson Plan.  138
C. Standardized assessment tool for transthoracic echocardiography (TTE).  143
D. Standardized assessment for clinical evaluation of echocardiography.  144
E. Institutional Review Board memorandum of approval.  145
F. Random sequence generator used to randomly assign participants to groups.  147
G. SPSS normality testing and subsequent parametric analysis of pilot data.  149
H. SPSS normality testing and parametric analysis of experimental data.  165
I. SPSS dependent sample t-tests for within-group comparisons, Round One.  176
J. SPSS dependent sample t-tests for within-group comparisons, Round Two.  179
K. SPSS ANCOVA analysis of pretest/posttest scores for Group I and II.  185
L. SPSS bivariate correlational analysis for simulator and clinical activities.  190
M. Photograph of UI of an actual ultrasound imaging machine.  201
N. Photographs of the UI for the simulator.  202

References  204
List of Figures

Figures

1: Switching replications experimental design with multiple assessments. 37
2: Application of the PC$^3$ instructional design to the educational unit. 43
3: R2D2 instructional design applied to the educational unit. 46
4: Line plot of mean scores by group versus assessment round. 79
5: Scatterplot of time to complete simulator views vs. round two scores. 87
6: Scatterplot of simulator combined score vs. clinical practicum score. 88
7: Scatterplot of simulator performance factor vs. clinical practicum score. 89
8: Scatterplot of simulator performance factor vs. clinical performance factor. 90
9: Simulated ultrasound view of the human heart. 100
10: Same view on an actual person using actual ultrasound equipment. 100
Chapter 1

Introduction

Background

The effect of formal integration of simulator technology into an existing curriculum in undergraduate level health care education was the subject of interest. Established instructional design (ID) techniques were used in the simulation integration, with a focus on the use of applied constructivist methods. Quantitative analysis of learning outcomes and correlational analysis for evidence of the transfer of the simulator-based learning outcomes to clinical skills was performed.

The host curriculum for the integration effort was contained within a health sciences bachelor’s-degree program for medical ultrasound, specifically teaching the use of ultrasound for examination of the human heart and vascular systems, and thus named cardiovascular sonography. The cardiovascular sonography degree program was offered at a private, not-for-profit university in the Southeastern United States, Nova Southeastern University (NSU), whose main campus is in Ft. Lauderdale, Florida. However, the cardiovascular sonography program is located at the Tampa Regional Campus, one of several operational locations of NSU. The existing curriculum course
that hosted the integration of the experimental instructional unit was itself contained within the Blackboard electronic learning management system (LMS). The target population was the bachelor's-level students enrolled in the cardiovascular sonography program. The experimental instructional unit incorporated the use of an already acquired high fidelity simulator for transthoracic echocardiography (TTE), transesophageal echocardiography (TEE) and cardiac anatomy known as HeartWorks, produced by Inventive Medical, Limited of London, England (www.inventivemedical.com). The HeartWorks simulator was utilized within a dedicated simulation lab, and in the ultrasound training lab as regularly used within the degree program. Two of the existing faculty of the program served as subject-matter-experts (SMEs), assisting with the delivery of the instructional unit, providing evaluative feedback on the effectiveness of the ID, and assisting with assessments of student participants’ performances.

The educational outcomes of the newly developed instructional unit were evaluated using quantitative parametric statistical analysis of standardized test scores, after confirmation of random distribution was performed. Baseline pre-testing and multiple post-experimental assessments using standardized tests were utilized. Correlational analysis of simulation-based educational outcomes to the actual performance of clinical skills was performed in an effort to validate knowledge transfer from simulator education into clinical practice.

A randomized, switching-replications experimental design was used (Trochim & Donnelly, 2008, p. 234). This design featured two rounds of experimentation and three
waves of assessment, with participants randomly assigned to an experimental and a control group who switched roles after the first round of experiment and assessment. The switching replications experimental design eliminates the need to deny a possibly beneficial treatment to the persons in the control group. This type of design assures that all participants receive the experimental treatment, thereby controlling for potentially powerful internal threats to validity (Trochim & Donnelly, 2008, p. 171) that will be discussed in further detail in a later section.

In the first round of the switching replications design, the control group did not receive the experimental treatment but did continue to receive existing course content by the normally used methods as in all previous years of the course. After the initial experimental group went through the instructional unit, they too continued to receive the existing course content via the same methods as in all previous years of the course. Assessment took place at baseline, upon completion of the first experimental round, and again after the second round of experimental intervention. The switching replications design came into play, when the previous control group performed the experimental treatment after the first round of testing; thus the groups replicated the experiment but switched roles (Trochim & Donnelly, 2008, p. 234).

There were a minimum of confounding factors present, as the same two faculty members have delivered the same standardized testing and content for the course that was the host for the experiment for the last three years, utilizing the same equipment, labs, simulator, faculty, and evaluative testing instruments as in the previous years. The only
major difference was the new experimental instructional unit that formally integrated the simulator use into the curriculum, versus the previous ad-hoc use of the technology.

**Problem Statement**

The problem was the attainment of meaningful integration of simulator technology into existing curricula. This has been recognized as a common issue with the use of simulator technology in health care education (McGaghie, Issenberg, Petrusa, & Scalese, 2010; Wittels, Takayesu, & Nadel, 2012; Motola, Devine, Chung, Sullivan, & Issenberg, 2013). The use of simulation technology in health care education has expanded rapidly, especially in the last 10 years as a response to multiple factors influencing the way health care education is delivered: continuing reductions and restrictions in available clinically-based learning environments, pressures to increase cost-effectiveness of instructional delivery (ID) and educational delivery, increasing professional standards, improvements in fidelity and decreases in cost of the technology itself, and other factors (Motola et al., 2013; Wittels et al., 2012). Many educational institutions that have acquired simulation technology have experienced difficulties in attaining meaningful increased learner and instructor satisfaction, improvements in learning outcomes, measurable transferable learning from the classroom to the clinical setting, and full utilization of the technology (Arthur, Levett-Jones, & Kable, 2013; Cook et al., 2013).
A tendency has been identified for simulator technology to be treated as an add-on to the curriculum with little effort to integrate its use fully into the learning environment (Masters, 2014). The generally-held positive outlook on the effectiveness of technology–enhanced simulation use in health care education is not well-supported by rigorous empirical research, specifically into areas of effective ID (McGaghie et al., 2010). Therefore, health professions educators and administrators have few or no guidelines for the design of effective ID systems utilizing simulation resources that are typically highly capital-intensive resources requiring viable evidence of reasonable return-on-investment. Furthermore, most educational research in this area involves simulation use in medical school settings for the training of physicians, with little emphasis on the training of allied-health practitioners (McGaghie et al., 2010).

**Dissertation Goal and Hypothesis**

The goal was to add to the body of knowledge of ID, a working template for the successful integration of advanced simulation technology into an existing health care education curriculum. The specific curriculum in this case was for cardiac sonography, with the integration utilizing established applied constructivist educational techniques that have proven effective in the development of other educational units. This addition to the body of knowledge will benefit both the theory and practice of ID, as future researchers in ID theory may find this example of a practical application of ID theory to a
working integration to be instructive, and practitioners may find the ID template used to be a useful model in various circumstances. Similarly, educators in health care programs who wish to integrate simulations technology into their own curricula may benefit from use of the ID template for their own integration efforts of simulation technology.

Statistical analysis of learning outcomes and correlation to clinical outcomes provided supporting evidence of the guiding research hypothesis: students who perform the experimental educational unit will achieve statistically significant higher scores on standardized assessments. If the findings did not support our research hypothesis, then the null hypothesis of: there was no statistically significant difference in standardized assessment scores between students who have received the experimental intervention and those who have not, would be supported.

Stated in an even more fundamental way, it was sought to determine if the formal integration of the use of the simulator technology into the existing course curricula via the constructivist-based instructional unit was worth the time, effort, and expense required. This was determined by the findings of statistically significant improvements in educational outcomes and a positive correlation to clinical outcomes, indicating knowledge transfer from use of the simulator to the real world. If no statistically significant improvements were seen in educational assessment outcomes over the informal approach previously used, then the need for formal integration of simulator use into existing curricula so strongly called-for in the literature would have been contradicted.
Research Questions

The following research questions were addressed:

1. What is the current state of the literature regarding the integration of simulation into existing curricula, including best-practices and identified gaps in the research regarding such implementations?

2. What are the foundations of the proposed experimental educational unit from both a learning theory and ID theory standpoint, and the applicable ID methods to be used to design and develop the instructional unit?

3. What are the resulting effects on learning outcomes that can be attributed to the experimental unit, as analyzed by quantitative methods?

4. What evidence is found supporting the existence of knowledge transfer from simulator skills to clinical skills?
Limitations

A primary limitation was the available sample size. The number of students in the population available at the time of proposal was $N = 23$. The small potential sample size was compensated for by the switching replications experimental design (Trochim & Donnelly, 2008, p. 234), as discussed further in Chapter 3. Additionally, any future retrospective analysis of standardized clinical skills test scores by students in previous classes in comparison to study participant scores would have a sample-size limitation, as there have been a limited number of students moving through the program in prior years, for a total of $N = 26$. Waiting for another entering class would not have changed this limitation, as the following class entering in May of 2016 matriculated 19 total students, and the incoming future class of May 2017 will likely be limited to 24. These overall numbers of future, current, and prior student cohorts are a limitation to both the current and possible future research efforts, and may affect the statistical generalizability of outcomes to other settings.

Limitations may have existed as to the male-to-female ratio of the participants, education level, prior exposure, or other demographic factors. These possible demographic influences were controlled-for through the random assignment of participants to either the experimental or control group. In addition, a limited but appropriate statistical analysis was used to identify any major demographic differences
between participants in the control versus experimental groups after randomized assignment, with none found.

Another fundamental limitation was the contextual setting for the research, which consists of an entry-level allied health undergraduate degree program, at a private, not-for-profit institution in the Southeastern United States. The contextual setting plus the probable demographics of the participants may make the findings only partially extensible to other contexts of higher education, outside of similarly structured allied health educational programs. However, the use of a randomized experimental design does improve the generalizability of the results (Trochim & Donnelly, 2008).

Another primary limitation was the time available to participate in the experimental educational unit by the participants. As there was only one HeartWorks simulator available, the logistics of scheduling sufficient time for each participant to perform the exercises and tasks with the simulator was a limitation. However, with careful management and improvements in controls suggested by the pilot run-through in the Winter of 2016, all participants were able to schedule the required time on the simulator.

Another time-related limitation that leads into the area of participant attitudes was that the students had a full course load in all of their semesters, including the semester in which the experimental unit finally took place. Although educational benefit was both anticipated and later seen by the participants, it was recognized that there was the possibility of perception that participation was merely extra work, which could confound
the benefits of the experimental educational unit. This possibility was counteracted by careful explanation and continual communication on the importance of the participants’ role and likely learning benefits.

As mentioned in the discussion of participants’ time above, only one HeartWorks simulator was available in the targeted environment. Efforts to secure a second simulator in time for the research project did not come to fruition. The availability of this resource was a logistically limiting factor. However, this was handled through careful scheduling of access to the simulator for all participants involved, as already mentioned. A weekly schedule with available times for each day was created and physically posted in the simulator center, and replicated on the Blackboard course supporting this effort. This detailed scheduling of participants’ time on the simulator minimized possible negative influence on the participants’ performance of the educational unit.

Definitions and Acronyms

The following definitions and acronyms are specific to this educational and research context, and are included here for the benefit of the reader:

**AMEE.** The Association for Medical Education in Europe, is an international organization with a presence in 90 countries that promotes excellence in the undergraduate, graduate, and continuing medical education of health care professionals.
The AMEE sponsors a number of research initiatives on a regular basis investigating various areas in medical education (AMEE, 2015).

**CLT.** In the context of educational learning theories, CLT stands for Cognitive Learning Theory, which views learning as the result of the processing of information in order to construct knowledge and cognitive structures (Schunk, 2012, p. 490).

**Compensatory rivalry.** A social interaction threat to internal validity, occurring when participants in a control group become competitive with the experimental group, and try harder in response; this threat if present will result in an equalization of post-test performance between control and experimental groups, making a treatment effect more difficult to observe (Trochim & Donnelly, 2008, p. 171). This threat is controlled-for by the switching replications design (Trochim & Donnelly, 2008, p. 234).

**Diffusion or imitation of treatment.** A social interaction threat to internal validity, occurring when participants in a control group learn what the experimental group is doing and attempt to do the same thing in imitation; this threat if present will result in an equalization of post-test performance between control and experimental groups, making a treatment effect more difficult to observe (Trochim & Donnelly, 2008, p. 170). This threat is controlled-for by the switching replications design (Trochim & Donnelly, 2008, p. 234).

**Fidelity.** In the context of simulators and simulation technology, describes the relative level of realism portrayed. High-fidelity simulators/simulations portray real-life scenarios, situations, procedures, or processes very realistically, but usually at an
increased level of both cost and complexity. Low-fidelity simulators/simulations are only rudimentary representations of the real-life counterpart they portray, but are typically inexpensive and of low complexity and thus often cost-effective. (Scalese, Obeso, & Issenberg, 2007).

**ID.** Instructional Design is the practice of systematically planning for instruction where attention is given to nine related design elements: instructional problems, learner characteristics, task analysis, instructional objectives, content sequencing, instructional strategies, designing the message, development of instruction, and evaluation instruments (Morrison, Ross, Kalman, & Kemp, 2010, pp. 6-12).

**Medical simulator.** A model, representation, or device representing a patient or patient care setting, classified by degree of realism, (see: fidelity) components, method of presentation or modeling, and intended use. Medical simulators range from relatively simple task-specific training devices and partial-task trainers, through complex procedural trainers, computer-enhanced mannequins or computer-based simulations, and even virtual reality scenario-based fully immersive environments (Scalese et al., 2007).

**PBL.** PBL stands for Problem Based Learning, a collaborative discovery learning process where students work in groups on a problem or issue that may not have a singular correct solution, but instead may be resolved from multiple viewpoints or approaches. Instructors are facilitators who provide guidance and support but not answers to the problem being addressed (Schunk, 2012, pp. 65, 316).
PC³. In the context of ID the acronym PC³ represents the terms Problem, Cognitive, Conversation, and Collaboration, the elements of an ID model created by Jonassen (1999) in an effort to promote the practical use of constructivist learning theory in technology-aided learning environments in a manner meaningful to instructional designers.

Resentful demoralization. A social interaction threat to internal validity, occurring when participants in a control group become angry or demoralized and give up; this threat if present will result in an exaggerated positive treatment effect (Trochim & Donnelly, 2008, p. 171). This threat is controlled for by the switching replications design.

R2D2. In the context of ID the acronym R2D2 represents the terms Recursion, Reflection, Development, and Design. The R2D2 model of Colón, Taylor, & Willis (2000) allows the designer to continuously update and revise the instructional model based on feedback from a participatory design group and the collection of experiences gained from use via recursion and reflection.

SBHE. Simulation-based healthcare education, its increasing use driven by and characterized by a need for training methods not solely based on clinical apprenticeship, a need for practice within a controlled environment, a desire for standardized and available-on-demand educational opportunities, and an increased focus on patient safety (Motola et al., 2013).

Scaffolding. In the context of constructivist epistemology, scaffolding is a process of selectively providing support to a learner to allow the mastery of key
knowledge or features. The instructor controls task elements that are beyond the learner’s abilities at that point, thereby extending the effective range of the learner and permitting the attainment of tasks or knowledge that would otherwise not have been possible. As the learner progresses in capability, the instructional support is selectively diminished or withdrawn (Schunk, 2012, pp. 245-246).

**Sonographer.** A sonographer is a highly trained allied health professional, skilled and knowledgeable in the use of medical ultrasound for the diagnostic imaging of the human body. A sonographer is qualified by professional credentialing, academic training and education, and clinical experience to provide diagnostic patient care services using ultrasound and related diagnostic procedures. Sonographers are responsible for the independent operation of sonographic equipment, and for performing and communicating the results of diagnostic examinations using sonography. A sonographer uses her cognitive ability to identify, record, and adapt procedures as appropriate to anatomical, pathological, diagnostic information and images; uses independent judgment during the sonographic exam to accurately differentiate between normal and pathologic findings, and analyses sonograms, synthesizes sonographic information and medical history, and communicates findings to the appropriate physician (Society of Diagnostic Medical Sonography, 2015).

**Transesophageal echocardiography (TEE).** TEE is the medical ultrasound imaging of the heart performed by the introduction of a flexible remotely guided ultrasound transducer into the esophagus after mild sedation of the patient. Advantages
include the positioning of the transducer in near proximity to the heart without interference from other anatomical structures; in theory this should allow for greater ease in obtaining diagnostic imagery of heart anatomy. However, disadvantages include the relative difficulty of remote transducer manipulation, and the invasive nature of the examination. A high level of training and experience is needed to properly perform this type of exam. TEE is performed exclusively by physicians, usually anesthesiologists or cardiac specialists, due to its invasive nature (Chetlin et al., 2003).

**Transthoracic echocardiography (TTE).** Medical ultrasound imaging of the heart performed via an ultrasound transducer placed on the external surface of the human body (thorax) by maneuvering around the rib cage, sternum, and clavicle to view the requisite anatomical features of the human heart. A high level of skill is required to properly perform these examinations, although TTE is routinely performed by non-physician allied health specialists in medical sonography i.e. sonographers (Chetlin et al., 2003) throughout the United States, Canada, and the United Kingdom.

**ZPD.** Zone of Proximal Development, defined as the amount of learning possible by a learner given the optimum conditions for learning (Schunk, 2012, p. 500).
Chapter 2

Review of the Literature

The review explored the current state of the literature addressing the integration of simulation technology into existing curricula. Areas of investigation included the use of simulations in undergraduate and health care education, and its integration into existing curricula. Additional reporting included the use of established ID methods in simulation integration efforts, with a further focus on the use of constructivist methods and techniques.

Four sections address different aspects of the integration of simulation technology into existing curricula. First is an exploration of the use of simulations in undergraduate and healthcare education. The second section discusses the integration of simulation into a healthcare education curriculum. The third section describes literature findings involving the use of ID practices and simulation technology. The final section discusses the practical application of constructivist design in education, particularly involving the use of simulation technology. Each area of exploration is discussed separately and then in their entirety as a group in a summary of the literature review, discussing themes identified in the literature. Particular note was made of methods, outcomes, and
suggestions for further research as reported by the various authors, and expanded upon as needed in Chapter 3.

**Simulation Use in Undergraduate and Healthcare Education**

A set of guidelines on the use of simulation technology in healthcare education is found in the work of Motola et al., (2013), whose exhaustive review of the current literature resulted in a best evidence practical guide sponsored by the Association for Medical Education in Europe (AMEE). Guidelines and recommendations on simulation use in healthcare education contained in this practical guide include the following key points. Simulation-based healthcare education (SBHE) is usually undertaken as an additional resource to complement or strengthen an existing curriculum. As such, it is important to determine the desired educational outcomes of simulator use, and use these desired outcomes to guide the integration of simulation technology into the curriculum. Curriculum integration is stated to be the key to the successful use of simulation technology, and must be supported by administrative entities in the form of faculty training, time and resource allocation, and technical support. Successful integrations follow three phases; planning, implementation, and evaluation, in a continuous cycle of revision as needed. Feedback, deliberate practice, individualized learning, and team training are all critical aspects of the use of simulations in healthcare education; per Motola et al. (2013) all are supportive of constructivist learning principles. Mastery
learning and capturing clinical variation are critical aspects of healthcare education that simulation can address effectively. Increasing limitations on the availability of traditional clinical learning experiences will cause an increased dependence on the substitution of SBHE. Recommendations for future investigation include a focus on optimizing ID in the use of simulations, outcomes-based measurement of the results of simulations-based education, and a shift in focus to translational outcomes, i.e. results from the use of simulation in the educational setting transfer directly to improved clinical practice, in turn resulting in improvements in patient outcomes (Motola et al., 2013).

Similar guidelines are found in the work of Rutten, van-Jooligen, and van-der-Veen (2012), who address the following questions on the learning effects of computer simulations in undergraduate science education: how the use of computer-based simulations can enhance traditional educational models; and how computer-based simulations are best used in order to improve learning processes and outcomes. Strong evidence was found supporting the use of simulations to enhance traditional instruction. An analysis of scaffolding techniques is provided, that are stated to support the scientific discovery learning process, and classified as instructional support to the learner. Per Rutten et al. (2012) scaffolding is an instructional technique that is consistent with constructivist learning theory, who noted the importance of effective lesson planning, realistic scenario design, and most importantly, integration of the simulator use into the curriculum. Large effect size variations were seen to be obtained by changing the levels of instructional support or the design and integration of the simulator into the curriculum,
thus pointing out the critical nature of these factors in the effective use of the technology. Future work exploring the role of instructors and their own training on use of the simulator systems was called for, as well as further investigation into the integration of simulation into the overall curriculum.

Nuzhat, Salem, Al-Shehri, and Al-Hamdan (2014) explored the role of simulation use in undergraduate curricula, and the challenges facing such implementations. One of their outcomes was the observation that simulation use can provide for repetitive learning (an important constructivist concept) imparting needed skills and knowledge, while at the same time not placing patients at risk. This outcome was seen as a major benefit to the use of simulation prior to or in some cases in lieu-of traditional clinical practice. Simulation use was found to be effective at teaching procedural, diagnostic, and communication skills, and in improving student self-confidence. Student perceptions and acceptance of simulator technology are generally positive, with low-fidelity simulations the least favored, and high-fidelity simulations or task-trainers most favored and most readily accepted as evaluation tools. Institutional support in the form of staff, staff training, and resource availability of simulators (affecting individual duration of sessions) are important factors. Critical to the successful use of simulators are the student’s motivation to participate in simulator use, and the provision for adequate feedback by instructors. The careful planning and integration of simulator use into the undergraduate curriculum is noted as a critical process in the use of the technology, a conclusion shared
by the other literature analyzed in this area of inquiry as well (Motola et al., 2013; Nuzhat et al., 2014; Rutten et al., 2012; Wittels et al. 2012).

**Integration of Simulation into Healthcare Education Curriculum**

Analysis of the first area of the literature review on the use of simulator technology in undergraduate and health care education reveals the critical need for integration of the technology into curricula. The work of Wittels et al., (2012) describes the details of an integration effort into an existing health care education curriculum. Their inquiry involving an emergency medicine curriculum wherein simulation makes up 20% of the delivered teaching hours, reveals that learners rated their knowledge acquisition and clinical decision-making abilities to be improved after introduction of the integrated simulation curriculum. Further recommendations and observations on integrating simulation use into curricula include that simulation offers a risk-free method to acquire procedural and clinical knowledge at no risk to patients, thus complementing traditional instruction methods. The overall perceived learning experiences including those of traditional methods of case-study and lecture were rated by learners to be improved in the integrated curriculum. Wittels et al. (2012) propose that these effects were a result of the increased knowledge and increased clinical abilities afforded by the simulation use, and thus emphasizes the need for effective simulation integration. More-advanced students (seniors) reported less perceived effectiveness for simulation use, than
did beginner-level students (juniors). This was surmised to be due to the students’ own advancing clinical expertise, suggesting that simulation use should transition in the curriculum into an evaluative role, rather than a knowledge acquisition or teaching role, as learners advance in their learning.

Exploration of the integration of simulations into healthcare education curricula was performed by Arthur et al., (2013) whose modified Delphi survey sought to determine the most effective simulation design and teaching strategies for high-quality educational outcomes, designated as Quality Indicators for the design and implementation of simulation experiences. Through a three-round modified Delphi survey tool of thirty-two international experts in the use of human patient simulators in healthcare education, a set of Quality Indicator Statements was arrived at providing guidelines for simulator integration and use in undergraduate medical (nursing) curricula. These Indicators list a number of factors that are termed as pedagogical principles that Arthur, et al. (2013) posit are consistent with constructivist techniques, such as: scaffolding, progressive complexity, feedback or debriefing sessions, and group learning. Student and faculty preparation and training are seen as Quality Indicators as well. Emphasized is the importance of simulator integration into every possible course, aligned with curriculum goals and course objectives.

Similarly, the work of Masters (2014) on the integration of simulation into a baccalaureate-level nursing curriculum reveals that the best learning outcomes occur when the use of the simulation technology is integrated into the curriculum in meaningful
ways. It is emphasized that simulation should not be merely added-on to a curriculum as a new feature or tool, but instead needs to be fully integrated. Keys to successful simulation integration into the curriculum are: planning for comprehensive debriefing time after simulator use, extensive involvement of all faculty members involved in the use of simulation to include training time for the faculty themselves in simulator use, and a strong emphasis on student preparation. The importance of using established educational practices in the ID, integration, and implementation of simulation into the curriculum is strongly emphasized.

**Instructional Design in Healthcare Education Using Simulation**

Exploration of the literature involving ID in the use of simulation reveals the work of Chiniara et al., (2013), who created a taxonomy and conceptual framework for ID and media selection for the use of simulation in healthcare education. The resulting ID framework consists of four levels: the media (method of delivery), modality (type of simulation), instructional method, and presentation. A media and simulation modalities diagram is provided to assist in the selection of appropriate simulator technology based on the desired learning outcomes. It is important to align simulator use with learning objectives in the ID process. Emphasis is given to the relative lack of quality research on best-practices in ID for simulation use in healthcare education, with this taxonomy and framework proposed as a starting point for such efforts.
The relative lack of quality research in the area of ID in the use of simulation is illustrated by Schaefer et al., (2011). This literature review analyzed 4,189 research articles from 1999 to 2010 on the use of simulation in healthcare education, of which only 221 qualified as studies, 51 had a meaningful strength of findings statement, and 39 were found with significant translational outcomes, e.g. outcomes that translated from the didactic learning setting to the practical application setting, in this case clinical skills or practice. The analysis focused on four areas of the reviewed research: validation of the simulator (for its intended educational use), validation of performance evaluations, research design, and an analysis of translational impact (using the Kirkpatrick scale for likelihood of significant findings). An analysis of whether research was based on a stated or implied theory reveals that less than 25% stated an underlying educational theory. Another key finding was that researchers in this field should attend to established best practices in researching the effectiveness of simulator implementations. Included are a useful set of criteria to evaluate one’s own research papers or articles. The article ends with a statement that the research reviewed was of insufficient quality to establish meaningful best practices in ID and pedagogy, as was the original goal intended, and as indicated in the title of the article.

Further efforts to establish the current state of ID in the use of simulators in healthcare education comes from Nestel, Groom, Eikeland-Husebø, and O'Donnell (2011), who performed a meta-analysis of the applicable literature from 2000-2010, focusing on the use of simulations for learning and teaching medical procedures and methods, with 81
articles selected for analysis out of an initial 1,575 found. Although the technical quality and capability of simulations in general have increased dramatically in the decade included in this article, the focus was not on the technical criteria of simulations, but rather on reported outcomes from simulation use. Findings include that the use of simulation improves learner’s clinical knowledge and skills in the majority of cases, that both learners and instructors express high satisfaction with the use of simulator technology, and that few researchers focus on long-term gains. This article emphasizes that curricula with simulation use must focus on accessibility, transferability of skills to the clinical setting, and context-setting, and should be strongly based on established ID and educational theory. Thus, further work is recommended to optimize the alignment of learner, instructor, simulator, setting, and educational goals including long-term educational and clinical outcomes.

A focused effort to define the most effective ID features in the use of simulation technology in education is found in Cook et al., (2013), who compared the effectiveness of ID features via an exhaustive systematic review of the literature and meta-analysis of the findings. They sought evidence to identify the ID features considered to be most effective in the use of simulation-based education. Identified in order of greatest pooled-effect size noted, the most effective ID features are: range of difficulty, repetitive practice, distributed practice, cognitive interactivity, multiple learning strategies, individualized learning, mastery learning, feedback, longer time of use, and clinical variation. It is noted that many of these most-effective ID features are consistent with
constructivist techniques, especially those of cognitive interactivity, multiple learning strategies, individualized learning, and feedback. Cook et al., (2013) recommend that further work be done to evaluate combinations of these ID features, and for increased effect with specific learners and specific learning environments. Their conclusions include that future work should go beyond the simple presence of the most-effective ID features to explore variations in timing, delivery methods, and basis of use. Cook et al., (2013) provide the following list of six themes identified in the literature:

- Compare different approaches to grouping learners
- Compare different design features to enhance instructional effectiveness
- Compare different levels of instructor training or presence
- Evaluate the addition of one or more other learning modalities (e.g. lecture, Computer Aided Instruction (CAI), or other simulation) to baseline simulation training
- Evaluate the addition of simulation features or effects to enhance sensory experience
- Compare two technology-enhanced simulation modalities. (p. 875)

The closing recommendation suggests that future work in this area should consider the costs of simulation use in order to optimize cost-effectiveness: the best combination of features and modalities with the best learner and patient outcomes at the lowest cost.
Constructivist Design in Education

Vogel-Walcott, Gebrim, Bowers, Carper, and Nicholson (2011) compared ID features consistent with cognitive load theory (CLT) versus problem-based learning (PBL), a method supportive of constructivism. Addressed is how the extent the two educational theories are applied through ID affect the learner’s ability to acquire low level knowledge (procedural, declarative) and high-level knowledge (decision-making, conceptual, integrated). Findings suggest that there was no difference between the constructivist-based method of PBL versus a CLT approach (a process-oriented worked-example) for the acquisition of all knowledge types except one; the exception was that the integrated knowledge category was better-retained both initially and long-term by the subjects taught under the CLT method of a process-oriented worked-example. This finding was contrary to original expectations, and was attributed to be a possible result of sample size, unique population, setting, etc. However, the findings may have merit when considering the relative costs of implementing either type of educational model, constructivist approaches being deemed more time and resource-intensive. This was considered to be especially so when applied to novice learners, thus an educational situation with non-novice learners should be served equally well by either a constructivist or CLT approach.
Garcia and Pacheco (2013) reviewed a constructivist ID approach to mathematics education, on the effects of integrated constructivist-based simulation technology on learner attitudes toward learning. Learning techniques used included PBL, social and cooperative group learning, contextual knowledge, adaptive knowledge transfer, and extensive integrated feedback as part of the constructivist-based approach used. It was concluded that highly integrated constructivist-based educational units resulted in high levels of learner motivation, collaboration, and discussion; all viewed as positive educational outcomes improving the learner’s overall attitudes toward learning.

Similarly, Adamson (2010) examined the use of simulations and the learning effects of the constructivist design features used. He describes the transition of a traditional lecture-based curriculum into a blended format using simulation and constructivist-based learning principles, subject to the constraints of a complex learning topic and the perceived need to provide for learner autonomy with a minimum of instructor educational unit. Constructivist methods incorporated in the new blended learning format include the ability of learners to set their own tasks (individualized learning), to be active participants in their own learning, group learning, and self-reflection. PBL design, scaffolding techniques, and contextually valid learning were used, all methods considered to be highly supportive of constructivist techniques. Throughout the implementation of the new learning format, many technical issues were addressed with some only partially overcome, such as limitations in the allowed size of embedded videos, resulting in what was considered to be a compromise solution when
compared to the ideal structure of the simulation system originally envisioned. The resulting simulation implementation thus had limitations and some built-in design flaws in comparison to the initial design concept. For example, what was described as over-scaffolding or providing too much initial information to the learners, was seen as a design flaw as some of the initial delivery difficulties in the beginning of the new course were addressed. Additional issues were described as a sacrifice of realism in the simulation implementation, again attributed to technical limitations encountered. However, the final outcome was seen as positive with many indicators of increased learner involvement and satisfaction.

**Constructivism Discussion**

Basic guidelines on the use of constructivism in educational design are provided by the seminal work of Schunk (2012). Recommendations on constructivist design in education include the observation that constructivism focuses on the interactions of learners in educational situations in the acquisition of new knowledge. Constructivism is considered to be a learner-centric versus instructor-centered approach, and often referred to as situated learning or situated cognition. The goal of the constructivist instructor is to create and manage learning situations to optimize knowledge and skills acquisition by the learner through enriched experiences. Constructivism emphasizes widely integrated curricula with active learning opportunities for students. Per Brooks and Brooks (1999)
(as cited in Schunk, 2012, p. 261) the guiding principles of constructivist-based learning environments are to: Pose problems of emerging relevance to the learner; structure learning around primary concepts; value the learner’s points of view; adapt curriculum to address the learner’s suppositions; and assess learning within the context of teaching, i.e. authentic assessment. To accomplish these principles, the techniques of discovery learning, inquiry teaching, peer-assisted learning, group learning through discussions and debates, and reflective teaching can all be structured and used to enable constructivist learning principles.

The central premise of constructivism is that learners create (or construct) new understanding by actively building upon prior knowledge and experiences (Schunk, 2012, p. 231). Learners are said to create meaning as internal representations based upon their experiences, rather than acquiring meaning directly from external sources. Summarizing the assertions of Baviskar, Hartle, and Whitney (2009), and Brooks and Brooks (1999), essential features of the use of constructivism concepts in educational practice are that:

- Learning is characterized by *cognitively active* learners
- Learning should happen *in context* and should be structured around related themes or primary concepts
- New knowledge constructs are *built upon prior knowledge*
- New knowledge should be *applied* and *feedback provided*
- Learner *self-reflection* on the learning process is a key learning activity
Currently, there is no single constructivism theory of learning. Instead, many forms of constructivism are found in the literature. Good, Wandersee and St. Julien (1993) documented 15 distinct uses of the word in combination with other descriptors in the literature at that time.

**Key Theorists of Interest**

The work of Dr. Lev Vygotsky, a Belarusian developmental psychologist, is foundational to constructivism. Although his work occurred nearly a half-century earlier, Vygotsky was unknown in the West until the early 1960s (Chaiklin, 2003). One of the key themes of Vygotsky’s work is the Zone of Proximal Development (ZPD). The ZPD may be described as the difference between what the learner can do either with, or without assistance or guidance (Schunk, 2012). The process of overcoming that difference is how all knowledge is constructed and where cognitive development occurs (Vygotsky, 1978).

Dr. David H. Jonassen’s work is noteworthy, particularly in the application of constructivist principles in the digital age of learning. Jonassen’s research focused on constructivism and constructivist learning environments, learning theories, technology use, problem solving, PBL, and learner-centric ID (Jonassen, 1997; Jonassen, 2000; Jonassen, Strobel & Lee, 2005; Jonassen, 2005). Jonassen viewed computers and other technology as tools to extend the mind, with some of his research focused on the use of
computers and technology as mind-extenders, or what he terms Mindtools (Jonassen & Carr, 2000; Jonassen, 2006). Much of his early work dealt with constructivism and its application through ID (Jonassen, 1997; Jonassen, 1999) and promoted the use of fundamental constructivist concepts in practical ways, advocating the use of foundational theories within the ID context to facilitate the best learning experience possible.

Jonassen stated that in a constructivist learning environment, the learning activities the learner undertakes of exploration, articulation, and reflection, are supported by the instructors’ activities of modeling, coaching, and scaffolding. These activities performed by both the learner and the instructor support effective learning in constructivist learning environments (Jonassen, 1999, p. 231). In an effort to promote the practical use of constructivist learning theory in technology-aided learning environments in a manner meaningful to instructional designers, Jonassen created the PC³ ID model (Jonassen, 1999), that will be explored in more detail in Chapter 3, Methods.

Two Foundational Constructivist Concepts

Scaffolding and cognitive apprenticeship are two important and foundational constructivist concepts that are used in the experimental educational unit. Bliss, Askew, and Macrae (1996) discuss an extension of Vygotsky’s ZPD via the concept of scaffolding. Scaffolding is an educational technique that helps the learner close the gap in cognitive ability found in the ZPD. Scaffolding works by initially providing high
levels of support to the learner that progressively decrease in a planned manner as they accomplish increasingly difficult learning goals.

Scaffolding takes the learner from the realm of the known, toward the understanding of what is yet to be known (Karagiorgi & Symeou, 2005). Scaffolding makes the learning of complex or difficult concepts and tasks possible, that may otherwise be outside of the ability of the learner, and is key to the process of cognitive apprenticeship (Reiser, 2004). Collins (1991) describes cognitive apprenticeship as the transmission of expert knowledge to a novice in a gradual manner via specific processes: task or problem modeling or demonstration, provision of performance feedback, scaffolding via decreasing levels of assistance as the learner progresses that allows the learner to become increasingly autonomous, and mentoring by monitoring progress, evaluating performance, and helping overcome specific weaknesses.

**Application in the 21st Century**

Quintana et al., (2004) saw scaffolding as a key design feature in software created for the learning of complex science concepts, and as supportive of cognitive apprenticeship. Azevedo and Hadwin (2005) describe the use of scaffolding techniques in computer-based learning environments and other technologies such as simulations and Web-based learning systems, and discuss the design implications for the use of scaffolding to support self-regulated learning and metacognition. They contend that
scaffolding within well-designed learning software systems can provide for many of the same structure, guidance, coaching, and helpful hints that a human mentor would. Cook et al., (2013) warn us that one of the most common shortfalls of the use of simulations in education is the failure to integrate its effective use into the curriculum. Ramdass (2012) cautions that in the design and use of computer based educational tools (particularly game-based ones) we must ensure that the learner is actually acquiring scientific skills and knowledge, and not just getting better at using the tool itself (or playing the game).

**Summary**

This exploration of the literature on the use of simulations in undergraduate and healthcare education, and on the integration of such technology into existing curricula with a focus on ID techniques consistent with constructivism, has unearthed a few interesting themes. The most fundamental theme underlying most of the research reviewed is that simulation technology in education is effective in many settings and educational roles, but that more research must be done to determine the optimal use of this powerful learning technology. The call for further research was especially evident in the work of Chiniara et al., (2013); Cook et al., (2013); Nestel et al., (2011); Motola et al., (2013); Rutten et al., (2012); and Schaefer et al., (2011).

Themes identified are the need for adherence to and the use of established educational theory and ID in the implementation of simulation technology in healthcare
education (Arthur et al., 2013; Masters, 2014; Nestel et al., 2011; Schaefer et al., 2011), and a similar theme in the use of established research methodology when studying and reporting the results of the same (Arthur et al., 2013; Schaefer et al., 2011). Another important ID theme is the need for careful alignment to the desired learning goals and objectives in the any ID effort for the use of simulations (Arthur et al., 2013; Chiniara et al., 2013; Motola et al., 2013; Schaefer et al., 2011). A critical imperative is identified for the need to carefully integrate simulation use into the curriculum using established ID techniques (Arthur et al., 2013; Masters, 2014; Motola et al., 2013; Nuzhat et al., 2014; Rutten et al., 2012; Wittels, Takayesu & Nadel, 2012) and not just add simulation onto a course or curriculum as an interesting diversion or optional tool for occasional use. The majority of the reviewed research supports the use of constructivist-based educational techniques in the use of simulation technology in health care education (Adamson, 2010; Arthur et al., 2013; Cook et al., 2013; Garcia and Pacheco, 2013; Motola et al., 2013; Rutten et al., 2012). However, a cautionary theme was also identified that constructivism, though an effective approach to health care education and ID for the same, may not be the best approach, or perhaps, should not be the only approach to a given educational delivery situation in a specific setting. This was especially seen in the work of Vogel-Walcutt et al., (2011), and in the outcome of the research by Adamson (2010).
Chapter 3

Methodology

It was anticipated that there would be statistically significant positive benefits from the use of the simulator technology within the experimental educational unit. The experimental instructional unit formally integrated the use of a high-fidelity simulator into an existing curriculum, using selected ID to implement a practical application of constructivist learning theory. Per Issenberg, McGaghie, Petrusa, Gordon, and Scalese, (2005) and others, integration of simulation technology into the curriculum is critical to its effective use. Positive findings of statistically significant improvements in educational outcomes and positive correlation to clinical performance would support the formal integration of the simulator technology and justify the needed additional time, effort, and expense. Such findings would provide evidence that both effective learning and important knowledge transfer to clinical skills did occur from the use of the simulator within the experimental educational ID unit.

The minimum level of integration where positive effects may still be found was explored. The instructional plan as implemented pursued a focused, minimalist approach to the integration effort, as discussed in later sections. This focused approach occurred in response to strong time and availability constraints resulting in feasibility issues.
However, the resulting more-narrowly-focused experimental educational unit used still contained all of the initially designed, key constructivist ID features.

The experimental educational unit ID used scaffolding techniques and cognitive apprenticeship in the educational use of a high-fidelity human heart and TTE/TEE simulator. The simulator, known as HeartWorks (http://www.heartworks.me.uk), comes with rudimentary learning features that were incorporated into the experimental ID to develop a self-guided learning program that formally integrated the use of the simulator into the existing curriculum. The experimental educational unit used the simulator to teach basics of human heart anatomy and its imagery via medical ultrasound, and was undertaken only after the student participants were first trained on the use of the simulator itself to avoid any pitfalls due to learner unfamiliarity with the system, addressing the issues brought forth by Ramdass (2012).

**Research Design**

A switching replications experimental design was used (Trochim & Donnelly, 2008, p. 234) with random assignment of participants to a control group and an experimental group, with three waves of testing. Testing including a secondary assessment of clinical skills. The switching replications design is shown in the experimental design diagram in Figure 1. The randomly assigned control and experimental groups are shown receiving an initial pre-test, then receiving the treatment
or not, followed by a post-test. Then a replication of the treatment occurs with what was previously the control group, who after the first experimental phase switch roles and become another experimental group. A final post-test is performed, including a secondary assessment. Standard experimental design notation is used such as described by Gay, Mills, and Airasian (2011, p. 265), and by Trochim and Donnelly (2008, pp. 205, 234).

<table>
<thead>
<tr>
<th></th>
<th>R₁</th>
<th>O</th>
<th>X</th>
<th>O</th>
<th>O</th>
<th>O₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₂</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>O₂</td>
</tr>
</tbody>
</table>

Key: R₁ = randomly assigned group, O = observation or test, X = treatment, O₂ = secondary observation or test

**Figure 1: Switching replications experimental design with multiple testing.**

The switching replications experimental design was chosen since students are the source of volunteer participants, and there was a concern that those assigned to a control group may be denied possible positive benefits of the treatment. In this case, the treatment is the educational unit involving use of the simulator to learn the anatomy of the heart and practice obtaining its standard medical views via ultrasound. This is similar to the treatment concerns raised by Pesiridis, Sourtzi, Galanis, and Kalokairinou (2015) for example, and is addressed through the use of the switching replications design. Pesiridis et al. (2015) discuss that when a treatment is desirable or likely to be beneficial,
and when performed in a setting which may not be fully amenable to a no-treatment
group, random assignment of participants to an early treatment group and one that
initially serves as a control group, but will receive the treatment later may be desirable.
Thus the switching replications design overcomes a major problem of the typical
randomized controlled trial, since the need to deny the treatment (in this case, desirable
training) to the control group is eliminated.

Additionally, there can be very real, socially-based internal threats to validity that
may occur due to the perceptions of participants assigned to a control group who are not
isolated from the experimental group, but are socially connected (Gay, et al. p. 261).
These threats may arise from the perception that those receiving the treatment in an
experimental group are being afforded some disparate advantage over those in the control
group. As a result, behaviors in the control group such as compensatory rivalry, resentful
demoralization, and diffusion or imitation of treatment, may result (Trochim & Donnelly,
2008, p. 171). By ensuring that all participants receive the possible benefits of the
treatment the switching replications experimental design both alleviates investigator’s
ethical concerns over fairness issues and effectively controls for possibly very powerful
social threats to internal validity.

As an added practical benefit, since all participants receive the treatment and are
post-tested, the switching replications design provides additional data for analysis in the
form of a second set of before-and-after treatment scores. These scores can then be
examined in at least two ways; comparisons of the pre- and post-treatment test scores of
Group B and Group A; and the combined post-treatment scores of both groups in a correlational analysis for evidence of knowledge transfer of simulator skills to clinical skills. Possible outcomes to the analysis of the pre- and post-test scores of the separate groups when using the switching replications design include either a converge-diverge-reconverge pattern if only short-term gains are achieved as a result of the treatment; or a pattern of continual gain of the initial experimental group, if the treatment has long-term continuing effects (Trochim & Donnelly, 2008, p. 205). The hope was that the later would be observed, but the converge-diverge-reconverge pattern would prove to be just as powerful of an indicator of interventional effectiveness.

The available class of students was provided with informed consent and asked for their voluntary participation. Resulting volunteer participants were randomly assigned via random number generator to Group I, the initial experimental group, or Group II the initial control group. In the first phase of the switching replications design, after a baseline pre-test is given to all participants, Group I participants proceeded with the four week long instructional unit. Group B participants initially acted as a control group, but did receive traditionally-delivered course content as normally provided. After the initial phase of: pre-test of both groups, educational unit use by the experimental group, and post-test of both groups, the switching replications design comes into play in a second phase of the experiment, and the previous control group performs the treatment, followed by another post-test of both groups. This pattern of switching roles and replicating experimental participation, and multiple rounds of assessment are the key features of the
switching replications experimental design, with very real advantages that have been discussed above.

The normal didactic delivery of lectures, case studies, and lab instruction was still received by all participants, regardless of group assignment. Participants in the initial control group, and again after the initial experimental group participants switched roles after the first phase of the design, still received lecture and lab activities as has been the prior practice. This is important because it eliminates the shortcomings of an experiment comparing a treatment-versus-nothing. All participants including the control group in each phase of the design still received all of the other normally delivered learning opportunities to improve their knowledge and skills. This served to make the independent variable the formal use of the simulator within the experimental educational unit, rather than the effects of simply using the simulator versus not using the simulator, or of the validity of the use of the simulator itself. This is an important distinction, as the literature was seen to already contain many investigations into simulator versus no simulator comparisons, or simple simulator-use validation studies (Bick et al., 2013; Bose et al., 2011; Laschinger et al., 2008; Lammers et al., 2008; Scalese et al., 2007). This investigation explored the next level of inquiry beyond the simple question of the use of the simulator technology or not, instead exploring how the technology should be used.

In addition to direct educational outcomes, knowledge transfer to clinical outcomes was studied by the use of a separate assessment, a standardized hands-on examination of clinical abilities in echocardiography (see Appendix D). The scores from
this examination were analyzed for correlation to simulator-based skills as portrayed by post-test scores and time-to-complete factors. This exam is based upon and developed from the published guidelines of the American College of Cardiology/American Heart Association/American Society of Echocardiography Committee for the Clinical Application of Echocardiography (Cheitlin et al., 2003) and the American Society of Echocardiography (Mor-Avi et al., 2011). This was the same hands-on examination as used in the course curriculum over the past three years. Its use as a standardized assessment tool allows for the possibility of future research in the form of a retrospective analysis of scores attained by previous years’ students to the scores attained by participants, or even a longitudinal comparison to future classes of students.

Unfortunately, blinding of the participant groups to the experiment (e.g. placebo vs. treatment) was not possible, since it was obvious to participants that they were or were not performing the educational unit. However, the switching replications experimental design does address this, since all participants eventually performed the experimental educational unit (Trochim & Donnelly, 2008, p. 234). Nor was it possible to blind the subject matter experts used to evaluate pre- and post-testing and clinical outcomes, as they were involved in the scaffolding support of the participants while performing the educational unit. However, blinding was possible in the assessment of the educational and clinical outcomes; anonymization of the before-and-after tests prior to scoring and of the clinical outcomes evaluations was used, in order to control for the possibility of rater bias toward a particular participant.
The Experimental Educational Unit

The experimental educational unit used was a combination of two constructivist-based development models. After a review of the many published constructivist-based ID models, Jonassen’s (1999) PC³ development model, and the R2D2 development model of Colón et al., (2000) were selected. Instances were found of both models being used by others in learning situations, contexts, and environments similar to those extant in this investigation, with beneficial effects on learning outcomes as described below, and are thus considered to be proven instructional designs. Both models were used to develop the educational unit, as both bring different yet complementary constructivist design aspects desired in order to address the ID and development question.

The PC³ ID model was created by Jonassen in an effort to promote the practical use of constructivist learning theory in technology-aided learning environments in a manner meaningful to instructional designers (Jonassen, 1999). Jonassen is an early proponent of the adaptation of constructivist theory into ID, advocating the use of appropriate foundational theories within the ID context to facilitate the best learning experience possible (Jonassen, 1999). Jonassen’s PC³ model uses scaffolding concepts to guide students in the interpretation and resolution of learning problems. The Problem (P) or question is outlined, with student understanding ensured by supporting background information. Then, technology based Cognitive tools (C¹) help the learner engage and
interpret the Problem to enable the constructive learning experience. Such tools may include visualization and performance tools (e.g. simulations, virtual reality, modeling), knowledge modeling tools (e.g. mind- or concept-mapping), and information gathering or data mining tools. Additionally, Conversation \((C^2)\) and Collaboration \((C^3)\) tools are utilized to allow for the important social-learning aspects of constructivist theory, enable the co-construction of meaningful learning experiences related to the Problem.

Application of the \(PC^3\) model to the design of the educational unit is shown in Figure 2.

![Figure 2: Application of the \(PC^3\) instructional design to the educational unit.](image)
The conceptual framework provide by the use of the PC³ model was further enhanced via the use of the R2D2 ID model of Colón, Taylor, and Willis (2000). Both models were used in order to create a robust ID model consistent with multiple applied constructivist features in the ID aspects. The R2D2 model is itself based upon the foundational work of Willis (1995) who created an early ID model incorporating recursion and reflection as key components to the continual, iterative design process.

Willis asserted that there are three primary or first order principles of constructivist-based ID models; recursion, reflection, and participation (Willis, 2000). These features have been expanded and elaborated upon in the later R2D2 model. There are four underlying key themes within the R2D2 model: Recursion, Reflection, a non-linear iterative Development pathway, and the use of participatory Design. The R2D2 model allows the designer to continuously update and revise the instructional model based on feedback from the participatory group, and experiences gained from use. The R2D2 development model as applied to the design of this instructional unit is shown in Figure 3, below.
# R2D2 Instructional Design for Self-guided Simulator course

<table>
<thead>
<tr>
<th>R2D2 Focal Points</th>
<th>Explanation</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Define</strong></td>
<td>Begin with general ideas of the approaches to be taken during the design, allowing them to evolve throughout the design.</td>
<td>The course consists of four units, each focusing on a different aspect of the simulator use: 1) review of heart anatomy. 2) simulator-assisted practice of 6 standard imaging planes in dual view mode (anatomical and ultrasound views). 3) simulator-assisted practice of 6 standard imaging planes in ultrasound mode only. 4) comparative feature use to review and assess the ability of the student to capture the 6 standard planes, versus the ideal standard approach. The actual instructional unit design participatory group consists of faculty members combined with a focus group of second-year students of the degree program in which the simulator is utilized.</td>
</tr>
<tr>
<td>Create a participatory group who will be members of the iterative design team.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Design and Develop:**
<table>
<thead>
<tr>
<th>Preparation Tasks</th>
<th>Select development environment, media, and instructional strategies.</th>
<th>Microsoft PowerPoint to create PowerPoint multi-media presentation (hyperlinks, embedded pictures, videos).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Review and evaluate decisions.</td>
<td>Use of scaffolding techniques to build learner knowledge.</td>
</tr>
<tr>
<td></td>
<td>Update plans when necessary.</td>
<td>Participatory group review.</td>
</tr>
<tr>
<td>Creation</td>
<td>Create a prototype (surface characteristics, interface, scenario, hypertext, and instructional strategies).</td>
<td>Design</td>
</tr>
<tr>
<td></td>
<td>Evaluate design through expert appraisal and student feedback.</td>
<td>Create and capture illustrative images and video clips</td>
</tr>
<tr>
<td></td>
<td>Revise the prototype based upon comments and reevaluate.</td>
<td>Tie instruction to classroom discussions and hands-on labs.</td>
</tr>
<tr>
<td>Procedures</td>
<td>Demonstrate the natural progression of the classroom setting.</td>
<td>Evaluate recommendations for improvement and implement critical modifications.</td>
</tr>
<tr>
<td></td>
<td>Modify the progression to improve instruction.</td>
<td></td>
</tr>
<tr>
<td>Dissemination</td>
<td>Distribute the product when complete.</td>
<td>Post on Blackboard</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participatory group review and Continuous improvement cycle embedded in the program.</td>
</tr>
</tbody>
</table>

*Figure 3: R2D2 instructional design applied to an educational unit.*
**ID Conception: A Comprehensive Lesson Plan**

Combining Jonassen’s (1999) PC3 model and the Colón et al., (2000) R2D2 model resulted in a new unique ID framework, that in turn was used to develop the Comprehensive Lesson Plan for a 4-week long educational unit integrating the use of the simulator (Yoders, 2014). The Comprehensive Lesson Plan consisted of four one-week-long sub units, each week similar in structure but changing in content based on learning target (see Appendix A). The four sub-units progress through increasingly difficult learning outcomes: Week 1) Review of heart anatomy; Week 2) Simulator-assisted practice of six standard imaging planes in Dual View mode (anatomical and ultrasound views); Week 3) Simulator-assisted practice of six standard imaging planes in Ultrasound mode only; Week 4) The simulator’s Comparative feature was used to assess the student’s ability to capture the six standard images, versus simulator-generated versions of the ideal views.

As designed, supporting materials were to be made available throughout the instructional unit in a corresponding Blackboard course shell. Self-reflection was an important part of the lesson plan, facilitated by the use of a learner blog within Blackboard. Collaborative learning was facilitated via a group discussion board in Blackboard. Cognitive engagement was addressed by weekly Web Quest work. Assessment within the instructional unit was addressed through visually-based quizzes
for formative assessment of each sub-unit of instruction as well as a learner portfolio for summative assessment.

**ID Implementation: A Focused Lesson Plan, Later Revised**

After much consultation with project SME’s and repeated feasibility sessions including scheduling models and discussion of logistical concerns, a focused lesson plan was devised. Based on the prior comprehensive plan, the Focused Plan reduced the time and content delivery of the original plan in order to address logistical time constraints for participants, and their ability to access the single simulator system available (see Appendix B for the Focused Plan). The new Focused Plan was three weeks in duration with fewer learner contact hours, simplifying their workload as participants.

After the pilot run-through described in detail in a later section, additional adjustments were made to the educational unit plan. The now revised Focused Plan went back to a four-week duration instead of a three-week duration, as the consensus of the initial pilot volunteers was that three weeks may not be enough time to work with the simulator. An additional change occurred in one of the ways that participants used the simulator. Feedback from the pilot volunteers indicated that using the simulator in ultrasound-only mode to practice obtaining the views tended to be confusing, unless they could switch back and forth at-will between the ultrasound-only mode and the side-by-side anatomical and ultrasound mode. This was incorporated into the new educational
unit for the final actual experiment, with all participants trained on how to do the view-mode switching, and instructions for same added to the shortcut cheat-sheets and other supporting materials provided.

The experimental design for the implementation of the new, focused educational unit remained the same, still using the switching replications design (Trochim & Donnelly, 2008, p. 234) as described earlier. The switching replications experimental design consisted of a baseline assessment (pre-test) followed by the first cycle of the now four-week lesson plan performed by the experimental group; a second round of testing (post-test), followed by a repeat of the four-week lesson plan with the previous control group in the switching-replications experimental design, and concluded with a final post-test assessment and an additional hands-on clinical skills assessment.

Assessment Within the Instructional Unit

Both objective and subjective assessments were to be used within the instructional unit as originally designed. Objective formative and summative assessments were in the form of quizzes and tests. Subjective assessments were designed into the original instructional unit as summative evaluations of a learner-produced portfolio containing their best-works in the form of saved images, learner participation in Discussion postings, and the content of learner self-reflective blogs. An integral part of the instructional unit was the consideration of knowledge mastery for each sub-unit (weekly topic). This takes
the form of a joint discussion between instructor and learner, with the opportunity for the learner to increase the sub-unit length by up to two days, or proactively advance to the next topic.

**Quizzes and Tests.** Weekly quizzes consisted of a series of images corresponding to the weekly topic practiced on the simulator. The learner must correctly identify the subject of the image (view) and the key structural anatomy (SA) seen in each standard view. Summative tests at the end of each experimental phase collected all images and SA features into one comprehensive exam. All assessments were based upon the publicly available published guidelines of the American College of Cardiology/American Heart Association/American Society of Echocardiography Committee for the Clinical Application of Echocardiography (Cheitlin et al., 2003) and on the standardized echocardiography assessment tool as described by Bick, et al., (2013) with an example provided in Appendix C. Copies of this assessment were also provided to the participants as a tool to help them track their own learning progress.

**Portfolio.** Portfolios can provide a means of assessment of learner accomplishment that can add more depth than using only objective assessments. The learners self-selected images they produced during their daily activities on the simulator for inclusion in the portfolio. This was designed to engage the learner in ownership of their learning process, by showcasing their best work through each sub-unit of instruction. The portfolio can thus reflect progression toward the goals of the instructional unit, and may be used for assessment accordingly, although they were not
assessed in this case. The portfolio dovetails with the learner Blog, the combination allowing for both visual and verbal self-reflection of learning progress.

**Web Quest.** The learners choose their favorite anatomical structure of the heart in the first week of the instructional unit. They then did Web-based research for factual and case study information involving that anatomical structure, and captured images portraying that structure throughout their exercises. The collections were organized into a PowerPoint presentation, with any found cases or research papers summarized and included as supportive materials. The Web quest was reviewed and discussed with the learner and may be assessed for a participation grade when the educational unit is embedded within an actual course; it was not assessed in this case.

**Discussion group.** The peer discussion group will facilitate interaction and social learning. Participation of the learner in the group may be assessed for a participation grade when the educational unit is embedded within an actual course; it was not assessed in this case.

**Blog.** The blog was an important part of the learning process, allowing for learner self-reflection. Blog entries may be reviewed and discussed with the learner, and may be assessed for a participation grade when the educational unit is embedded within an actual course; it was not assessed in this case.
Addressing the Research Questions

The research questions were addressed by performance of the following activities. For the first research question, *what is the current state of the literature regarding the integration of simulation into existing curricula, including best-practices and identified gaps in the research regarding such implementations*, the literature review served as the means to address the need for the information desired. The goal of the literature review was the discussion of the published efforts of others on the integration of advanced simulation technology into existing curricula through the application of established learning theory and ID theory. Specific literature focused on the use of applied constructivist learning theory to develop such instructional units, and on integration into healthcare related curricula was sought.

The second research question, *what are the foundations of the proposed educational unit from both a learning theory and ID theory standpoint, and the applicable ID methods to be used for the research*, was addressed by discussion of the selected aspects of applied constructivist learning theory, and of the selected found instances of ID. Searches were performed for ID used in the integration of simulation technology in healthcare education situations, and for instances of such design based on constructivist methods. The selected found instances are those of Jonassen (1999) and Colón et al., (2000) as determined through investigation into similarly applicable ID efforts.
The third research question, *what are the resulting effects on learning outcomes that can be attributed to the educational unit, as analyzed by quantitative methods*, was answered by application of known examples of analytical methods to compare standardized learning outcomes as assessed both with and without the effects of the educational unit. Both between-group and within-group analysis of performance measurements were used.

And finally, the fourth question, *what evidence is found supporting the existence of knowledge transfer from simulator skills to clinical skills* was addressed by the correlative analysis of post-test results for both groups to the performance of the same participants in a clinical setting where the same procedures as practiced on the simulator were performed in a clinical setting on actual persons. If a positive correlation was seen between simulator-based outcomes and the performance of clinical skills this would supporting the existence of such knowledge transfer.

**Instruments**

The primary data collection instrument that was used in conjunction with both the pre-test and post-test assessments is a checklist, based on the standardized echocardiography assessment tool as described by Bick, et al., (2013) combined with information from the published guidelines of the American College of Cardiology/American Heart Association/American Society of Echocardiography
Committee for the Clinical Application of Echocardiography (Cheitlin et al., 2003) and the American Society of Echocardiography (Mor-Avi et al., 2011). Please see Appendix C for an example of this instrument, the basic transthoracic echocardiography evaluation tool (BTTEET).

The secondary data collection instrument used was the standardized practical examination given to the students every year in the course during the Fall term. This examination is based upon the publicly available published guidelines of the American College of Cardiology/American Heart Association/American Society of Echocardiography Committee for the Clinical Application of Echocardiography (Cheitlin et al., 2003) and of the American Society of Echocardiography (Ehler, 2001; Mor-Avi et al., 2011). Please see Appendix D for an example of this practical examination.

Data Collection and Analysis

The educational outcomes of the newly developed instructional unit were initially evaluated in two ways: pre- and post-test quantitative comparisons of standardized test scores between groups, and a correlational analysis of educational outcomes on the simulator compared to actual clinical performance of examination skills, in order to validate knowledge-transfer from the simulator into clinical practice. Secondary analysis included comparison of post-educational unit scores again, after the second round of educational unit, wherein the previously used control group then performs the educational
The scores of the two groups post-educational unit should not be significantly different for the second round of clinical evaluation scores, unless long-term educational effects of the experimental unit are observed, as previously discussed (Trochim & Donnelly, 2008, p.205).

The randomization of participants into control and experimental groups provides for probabilistic equivalence between groups in experimental design. However, additional testing for demographic differences between the groups may be performed, similar to that done by Barsuk, McGaghie, Cohen, Balachandran, and Wayne, (2009). Investigation of possible demographic differences helps rule out the existence of confounding factors when considering outcomes. One of the ways to test for such demographic differences is the chi-square test of independence. The chi-square test of independence allows a series of nominal variables to be tested for their independence from each other (Terrell, 2012, pp. 291-317). In this case the nominal characteristics were the basic demographic criteria of the participants, male versus female, age bracket, and educational background. An analysis showing independence of demographic characterizations allows for any effects observed to be attributed to the influence of the treatment applied, versus the external factor of demographic differences.

Most of the statistical analysis work however, involved an extensive set of quantitative statistical analysis tests that were performed on the collected assessment data. Group to group comparisons for baseline, first round, and second round assessments were made utilizing the independent samples t-tests. Analysis of within-
group performances at baseline versus first round post-test, and first round versus second round scores was also done using dependent sample t-tests ala Barsuk et al., (2009). Correlation between simulator skills and the participants’ performance in real-life clinical outcomes, indicating knowledge transfer, was evaluated by the use of ANCOVA test.

**Resources**

Access was needed to a suitable number of participants, who were students enrolled in the undergraduate Health Sciences education program in cardiovascular sonography, as previously described. Although a population consisting of one’s own students is considered to be a vulnerable one, it was also the only valid population when seeking to determine the effectiveness on specific educational outcomes of a focused educational intervention embedded within the normal educational pathway of these same students. Every effort was made to decrease vulnerability of the students who did volunteer, beyond the full disclosure and approved IRB form provide to every student who was a potential participant. Such efforts included continuous repetition that their participation was voluntary, that they could withdraw at any time, and that their participation or non-participation would not affect their grades in any way in their normal curriculum.

The next most-critical resource was the simulator itself and the simulator suite, and sufficient availability (time) for participants to access the simulator. This was
followed by the supporting infrastructure of normally-used computers, classroom and training lab facilities of the program. The larger-context of needed resources included course and content delivery as hosted in the Blackboard LMS, the network and infrastructure required to provide access to the LMS for both students and faculty. Access to IBM SPSS statistical software and Microsoft Excel was required for data tabulation and analysis. Additionally, the current faculty engaged in delivery of the existing course and simulator use was needed to facilitate the delivery of the instructional unit and assist in the evaluation of learning outcomes as subject matter experts (SME’s). All of the above resources were currently available in-house from the outset, minus research access to the students as participants, which was based upon Internal Review Board (IRB) approval and their own willingness to volunteer to participate.

**Execution**

The instructional unit as described was initially designed to be carried out over a single 16-week term. The preparatory groundwork took place for the class of students matriculating in May of 2015 in the first academic term of the program, the Summer term of 2015. This preparation consisted of a basic training regimen in the use of the simulator itself, that all students \( N = 23 \) did complete, and provision of quick-reference guides and a “cheat sheet” for ready reference while using the simulator. The study was originally planned to occur the next semester, the Fall term of 2016 (calendar year 2015).
However, proposal approval and IRB clearance consumed the majority of the Fall term. IRB approval was sought and obtained in the late Fall of 2015; the letter of IRB approval dated November 17, 2015 is included as Appendix E. Therefore, the plan was adjusted to occur in the students’ third term of the program, the Winter term of 2016. The assessment tools were modified to reflect the students’ progress through the curriculum, who would now be in their third semester of echocardiography courses (Echo III) in the Winter term; the original tools revolved around beginners at an entry-level, acquiring the six most basic clinical views of the heart. Since the students were now at a more advance level, six more advanced views of the heart were selected to be congruent with the level of instruction, and the assessments and educational plans adapted accordingly. These six advance views would be a normal part of the third semester learning goals in any case, and were substituted for the beginners-level views originally targeted.

The entire class was informed of the research project, and volunteers sought and provided with the approved IRB informed consent forms on February 12th, 2016 during a lunchtime learning session for the Health Professions Division (HPD) Research Day. The students were asked to provide their response by the end of the following week, Friday February 19th, 2016. The plan was for those volunteering to be randomly assigned to either the initial control or experimental group, with the experimental educational unit (intervention) performed by the experimental group early in the Winter term, continuing for four weeks. Then, the educational unit was to be repeated with the
control group in the following four weeks, as described in the Research Design section previously and per the switching replications experimental design.

Perhaps prophetically, the study proposal outlined a timeline wherein “the year, term, and student cohort of participants was ultimately based on the overall timing of the dissertation proposal approval and IRB approval, but should not extend the experimental phase later than the Fall term of 2016 in any case”, and mentioned the possible use of the next class of students to enter the program: “. . . the incoming student cohort of May, 2016 would provide the next possible participants. . .” if needed.

Unfortunately, moving on to the incoming class of 2016 proved to be necessary. Out of the initial group of students who were the incoming class of May 2015, now one less due to an academic dismissal (N = 22) who were asked to volunteer for the study, only nine of the 22 students initially agreed to participate. This was considered to be an insufficient number of participants for the full experimental plan as designed, so after a two-week period, during which other efforts were made to recruit further volunteers to no avail, permission was sought from the dissertation chair and committee to delay the experimental phase, which was granted. The full study with the switching replications experimental design was then re-targeted to involve the incoming class of May of 2016 and occur in the Summer term of 2016. Ironically, the most common reason given by the abstaining students for not volunteering was the workload in their Research Methods course that term that apparently was taking up too much of their time and requiring extra work effort, and thus they did not wish to volunteer for actual research.
**Pilot run-through.** So as to not completely waste the willingness to participate of the nine volunteers, it was approved that these students could act as a proof-of-concept or pilot group to help fine-tune the educational unit procedures and assessment instruments for the future, larger study. A minimal experimental plan was used, with no pre- and post-test assessment. With the single group of all nine (and later, ten) volunteers as the experimental group, the switching replications design of the full study was not utilized for the pilot activities. A tenth student volunteered in the first week of the pilot activities, after she stopped in to see what the volunteer students were doing on the simulator, and decided she wanted to participate after all.

Despite being of limited scope, all other planned activities were performed during the pilot, starting the week of March 27th, 2016, and for the following three weeks, culminating with a hands-on clinical assessment in late April 2016 as described below. Key aspects of the planned educational unit were still employed including scaffolding of learner support for participants, assessment via interactive image-based quizzes, online discussion board and blog for social interaction, learning, and self-reflection, and the Web quest to provide extended learning opportunities. Despite the limited number of participants and the limited nature of assessment during the pilot, some valuable insights were obtained, that will be discussed in the initial section of the Results chapter.

From a practical viewpoint the pilot run did provide an opportunity to fine-tune the assessment tools, including additions to allow notation of the exam room used, time started and ended and elapsed time overall, and other minor but useful details. The run-
through suggested areas for procedural and logistical improvements, particularly in the
way learner support via scaffolding was handled. The pilot highlighted time management
concerns for us as instructors, issues that were acted upon by scheduling which
SME/investigator would cover the simulations lab on certain days. Overall the ten
volunteers did well on their clinical scores, and in comparison to the abstaining students,
showed statistically significant positive differences in some aspects of their performance,
although there were no differences in time performance criteria.

**Applying lessons learned and preparation.** The pilot run-through and data
collection ended the Winter 2016 term in late April of 2016. With the new incoming
class of May 2016 imminent, some minor adjustments per the lessons learned in the pilot
were completed. Fine-tuning took place in how the learner scaffolding support would be
managed and in the assessment worksheet, and to the sign-up process the students would
use to reserve their time slots on the simulator. The assessment tools were reverted to
the original six basic views, taking them back to a level suitable for beginners, as
originally planned.

Once the incoming class of May 2016 was set ($N = 19$), on Friday May 20th,
2016 during new student orientation week a lunchtime meeting was held with the
students. The final agenda item for the meeting was an announcement and description of
the study and a request for volunteers. The approved IRB disclosure forms were handed
out and gone over in detail, with each page projected on in-class video screens and read
through in their entirety. Then a descriptive briefing including a hands-on demonstration
of the simulator was held, followed by a Q&A session. The students were asked to take their time think about whether or not they wished to volunteer, but to hand in their signed IRB forms indicating their willingness to participate or not, by the end of the following week, on Friday May 27th. They were informed that baseline testing would take place the end of the following week, on June 3rd, 2016. Eventually every student except one volunteered, resulting in \( n = 18 \) participants. The sole hold-out said that she was concerned with the extra time commitment needed and chose to err on the side of caution by not participating.

As the students turned in their signed IRB forms, over the period May 20 through May 27, 2016 each form was sequentially numbered one through 18. The signed forms were then stored separately from all the other paperwork and artifacts generated, per IRB stipulations. On Friday, May 27th, the 18 participants were randomly assigned to the initial experimental group (Group I) or to the initial control group (Group II) via a random sequence generator (please see Appendix F) with their designated form number acting to assign them to one or the other of the groups. Descriptive analysis of demographics comparing factors between the two randomized groups revealed a similar distribution of age range, averaging 23-28 years old for both Groups; and for education level, with a factored analysis yielding a 1.44 average for Group I and 1.22 average for Group II, within a range of 1 equaling bachelor’s level and 2 equaling master’s level. Unfortunately, the planned chi-square analysis of demographic characteristics was not possible, as the number of occurrences of some values were less than five, a number
which does not work well with the chi-square test (Terrell, 2012, p. 212). This is a direct result of the small number of participants available and a limitation that will be discussed in more detail in the Conclusions section.

**The full experimental unit.** Baseline testing for knowledge of the six basic views of the heart and the ability to identify structural anatomy of the heart in these views was held on Friday June 3rd, 2016 for all 18 participants in both Groups. Scoring data was collected for later analysis, then the Group I students began the first of four concurrent weeks of simulator use per the experimental educational unit starting the week of Monday, June 6 through Saturday, June 11th, 2016. Group I students signed up for two one-hour-long or four 30-minute long simulator sessions per week (Saturday times were offered per multiple students’ request throughout both cycles of simulator use in the switching-replications design, but these Saturday time slots were never actually used by either group of students). The fourth and final week of simulator use for Group I participants was Monday June 7th through Saturday July 2nd, 2016. Since no-one signed up for simulator use on Saturday the 2nd, the first round of testing was completed on Friday July 1st, 2016 for both Group I and Group II participants, using the same assessment tools as the baseline test.

At this point the Group I students had experienced both the benefit of the normally-delivered educational curricula plus the simulator use within the educational plan, while Group II student participants had only received the regularly delivered materials and methods. The expectation was for some positive improvements to be seen
in Group II scores over their baseline scores, since they were still being taught via normal methods, but that the Group I scores would be significantly positively higher than both Group II scores and their own baseline performances. This is exactly the pattern seen after later data analysis that will be discussed in the Results section following this chapter.

Starting the week of July 5th through the 8th, the participants in Group II then performed the experimental educational plan and used the simulator for the following four weeks, through the week ending August 30th, 2016. During this four-week period, Group I participants did not use the simulator but did continue to receive all normally delivered instruction via regularly delivered materials and methods. This was per the switching replications experimental design previously described.

The final round of assessment testing for both Group I and Group II occurred on August 2nd, 2016. The same assessment tools used in both the June 3rd baseline testing and in the Round One testing on July 1st, were used in the testing on August 2nd. This was followed by hands-on clinical assessment on Friday, August 5th to assess all participants’ ability to obtain the same six views in a clinical setting on an actual person, per the plan. Separate morning and afternoon assessment sessions were held on both August 2nd and August 5th in order to accommodate all participants in either the simulator lab or the hands-on lab, respectively. These sessions were organized by the students’ regular course lab group assignments, not by their participation in either Group I or Group II, resulting in a mix of experimental Group members in each testing session.
An additional source of data for correlational analysis and possible identification of long-term learning effects was the midterm exam for all students during the Fall term, held on October 27th and 28th, 2016. This midterm hands-on exam assessed the students’ skills in acquiring the same six basic views as those targeted during the experimental phase, thus providing another possible point of comparison.

All data was collected for later analysis, and after each session, an informal focus-group discussion was held seeking comments and feedback from participants. All participants were then given their choice of a $20 gift card from various establishments (Starbucks, Subway, Cold Stone Ice Cream, or movie tickets) that were paid for out-of-pocket as a personal thank-you and delivered to them by the last day of class for the Summer term, Friday August 12th, 2016. Data organization and statistical analysis was begun the week of August 15th, 2016. After numerous interruptions in the analysis work for: finals week, graduation ceremonies, an annual Departmental retreat, closure of the NSU Tampa campus due to the approach of Tropical Storm Hermine, an annual College-wide meeting, attendance at a national professional conference, and closure of all NSU campuses due to Hurricane Matthew, data analysis work finally recommenced in earnest in late October, 2016, culminating in January, 2017.
Summary

The experimental instructional unit was a four-week long, participant self-guided implementation of simulator technology, that was repeated in a switching-replications experimental design (Trochim & Donnelly, 2008). Important aspects of the ID used for the experimental instructional unit strongly adhere to established constructivist learning theory and practice, an overall design goal of the project. Use of this instructional unit was anticipated to have significant positive effects on the learning outcomes of participants. Expected were positive correlations to participants’ clinical performance. Such positive correlation would be evidence of effective knowledge transfer from the realm of simulator skills to the real world. It was anticipated that the results would uphold the desirability of formal integration, as evidenced by statistically significant positive educational outcomes and positive correlation of simulator-based skills to actual clinical skills. The working example thus validated should prove to be of benefit to both the theory and practice of ID, and to similar educational programs wishing to effectively utilize simulator technology.
Chapter 4

Results

The following results are from the initial pilot run-through and the full experiment using the switching replications experimental design. The limited outcomes of the pilot run-through will be briefly described first. In the full experiment, statistically significant positive effects for use of the simulator as part of the structured educational plan were found when assessing participants’ ability to identify and apply their knowledge of cardiac anatomical structures and of the six basic ultrasound views of the heart.

These positive effects were found in both between-group and within-group comparisons of baseline versus post-intervention performances, for both groups of participants. Controlling for the co-variant of pre-test (baseline) scores resulted in statistically significant positive results after performance of the educational unit. Only mild correlations were found between participants’ performance on the simulator and their ability to perform the same skillset in a clinical setting. That is, performance on the
simulator was only a mild predictor or indicator of performance when using actual medical ultrasound imaging equipment on an actual person.

This finding was contrary to initial expectations. After further consideration however this did make sense due to factors that will be discussed in detail in the following chapter in the Conclusions section. Please see Appendices H through L for detailed statistical output from SPSS in support of the following sections.

**Limited Results of the Pilot Run-through**

Important feedback from the ten participants in the pilot included that they were already under considerable pressure for time utilization, and since it was “not for a grade” they did not feel that the web quest and portfolio were a useful expenditure of their time. Even though some could see the educational value of these activities, all expressed that they felt that it was “extra work”. When asked if they might change their mind about the usefulness of performing these activities if they were graded as an integral part of a course instead of a volunteer activity, the universal response was that yes, they would then pursue those activities. This was an interesting outcome of the pilot that portrayed an unfortunate negative attitude even amongst volunteers toward the performance of learning activities unless they are graded. Doing an activity even if perceived as beneficial to their learning was not valued unless graded; learning for learning’s sake
seemingly did not exist, an outlook that forecast and foreshadowed similar outcomes in the later, full experimental run through.

Feedback from the pilot volunteers included insight into their low participation in the blog for leaner self-reflection. The consensus was that it was a waste of their limited time since it was not a graded activity. Only one out of the ten volunteers ever initiated a blog entry, despite multiple verbal reminders from the SMEs and myself during simulator activities, and additional emailed reminders. The use of the discussion board met with only slightly better participation, with three participants providing commentary to the instructors and some limited interaction with each other.

Another item of interest came about as the pilot run-through was wrapping up. The final activity of the pilot was the clinical assessment that was performed in the lab. This provided an indicator of how well simulator skills translated to the performance of those same skills on an actual person. Prior to this hands-on exam, that is a normally-given part of the Echo III course, about half of the 12 students who abstained from participating in the simulator training reportedly got together and decided that they were going to compensate for their non-participation. Two of these students told one of the SME/professors that they were going to “show you that we can do just as good as those guys” and rallied support from four other students to try to out-perform the ten volunteers in the hands-on clinical test.

This was a perfect example of a possibly confounding social interaction threat to the internal validity of a study, termed compensatory rivalry, previously defined and
discussed. In this case the non-participants became competitive with the experimental group who were seen as receiving special treatment, and tried to out-perform them during testing. This can result in an equalization of post-treatment performance between control and experimental groups, making a possible effect more difficult to observe (Trochim & Donnelly, 2008, p. 171). This threat is controlled-for by the switching replications design (Trochim & Donnelly, 2008, p. 234) as used in the full experimental study. However, as already described the switching replications design was not used for the pilot run-through.

Only an after-treatment assessment of a sort was performed, in conjunction with a normally-given hands-on clinical skills test late in the term, on Wednesday April 20th and Friday April 22nd, 2016. This was a hands-on skills assessment that would have been given to all students in any case. The scoring for all students, both the volunteers and those who abstained, was completed as always with the existing grading rubric to determine their grade for the clinical skills test. All of their performances were evaluated a second time with the experimental assessment tool. The balance of the students in the class who abstained from volunteering (now $n = 12$) thus acted as a de-facto control group.

The data from the pilot were tested for normality of distribution. A relatively normal distribution of all scores was seen via all three normality of distribution testing methods: numerical (skewness, kurtosis), statistical (Shapiro-Wilk), and visual (Histogram, Normal Q-Q Plot). Since normality of distribution was satisfied with
consistent results from both objective (numerical, statistical) and subjective (visual) methods, parametric statistical testing may be used (Park, 2008, p. 36). Please see Appendix G for normality testing and the subsequent parametric statistical analysis of the pilot data.

With a relatively normal distribution of the data and comparable mean values the assumption that the data are normally distributed is satisfied and parametric inferential statistical testing can be used (Terrell, 2012, p. 114). The parametric statistical test, independent sample $t$–test was performed to look for any significant difference in the clinical assessment metrics from the pilot. These tests yielded interesting and unexpected results. As expected, on average the volunteer students had significantly higher combined clinical assessment scores than the abstaining students with volunteers averaging 27 out of 33 possible points, and the abstaining students averaging 20.08; $t (20) = 3.241, p = .0015$ at $\alpha = .05$. When expressed as a percentage score the volunteers had significantly higher combined clinical assessment percent scores than the abstaining students, with volunteers averaging 82% versus abstainers averaging 61%; $t (20) = 3.239, p = .0015$ at $\alpha = .05$. The volunteer students had a significantly higher number of correct views in the clinical assessment versus the abstaining students, averaging 4.5 correct views out of 6 versus 3.33 for the abstaining students; $t (20) = 1.999, p = .0295$ at $\alpha = .05$. Volunteers had significantly higher average structural anatomy (SA) scores, averaging 22.50 out of a possible 27 versus 16.75 for the abstainers; $t (20) = 3.498, p = .001$ at $\alpha = .05$. 
Interestingly and unexpectedly there was no significant difference between the two student groups in total time to complete the six views, or in the average time to acquire each view. On average there was no significant difference in the total time to complete the six clinical views between the volunteers, who averaged 11.19 minutes, versus the abstainers who averaged 11.33 minutes; \( t(20) = -0.072, p = .943 \) at \( \alpha = .05 \). Similarly, there was no significant difference between groups in the time to obtain each of the six clinical views, with the volunteers averaging 1.86 minutes per view and the abstainers, 1.89 minutes per view; \( t(20) = -0.072, p = .943 \) at \( \alpha = .05 \). This statistical equality in time metrics may have been a result of the compensatory rivalry that took place, leading the abstaining students to try to be “just as good” as the volunteer students, by being just as fast. The compensating abstainers managed to match the speed of the volunteers. However, that speed may have been achieved at the expense of accuracy in their ability to obtain the views correctly and in properly identifying structural anatomy, as seen in the statistically significant differences in those metrics as already described above (please see Appendix G for detailed normality of distribution testing and subsequent parametric statistical analysis via SPSS of the pilot data).

**Full Experiment: Baseline Assessment Analysis Between Groups**

Initial data collection and descriptive statistics of baseline scores revealed comparable means between groups. The mean score for Group I was 27.88, and for
Group II, 29.22, expressed as a percentage score. A relatively normal distribution of the baseline scores for both groups was seen via all three normality of distribution testing methods: numerical (skewness, kurtosis), statistical (Shapiro-Wilk), and visual (Histogram, Normal Q-Q Plot). Since normality of distribution was satisfied with consistent results from both objective (numerical, statistical) and subjective (visual) methods, parametric testing may be used (Park, 2008, p. 36). Please see Appendix H for detailed normality testing and subsequent parametric statistical analysis. With a relatively normal distribution of the data and comparable mean values the assumption that the data are normally distributed is satisfied and parametric inferential statistical testing can be used (Terrell, 2012, p. 114).

The parametric statistical test, independent sample $t$–test was performed to look for any significant difference in baseline assessment scores between the students randomly assigned to Group I versus the students randomly assigned to Group II. Test hypothesis (two-tailed, non-directional): There was a significant difference in baseline assessment scores between the students assigned to Group I and those assigned to Group II. Null test hypothesis: There was no significant difference in baseline assessment scores between Groups (expected result).

The computed $p$ value (Sig. 2-tailed) of .830 is much greater than the alpha value of .05 divided by two, equals .025, for the two-tailed or non-directional test hypothesis. Therefore, the decision must be to not reject the null hypothesis (fail to reject). The critical value of $t$ for $df = 16$ at $\alpha = .025$ is 2.120 (+, -). With the computed value of $t$ at -
.218 well within the range of the critical value of \( t \) at + or -2.120, we must fail to reject the null hypothesis that there was no significant difference in baseline scores between groups.

The mean baseline assessment score for Group I was 27.89, for Group II 29.22, which are roughly comparable, and the computation of Cohen’s delta resulting in a very small effect size of .102, i.e. there is very little effect on the dependent variable, baseline scores, by the independent variable group assignment (Cohen, 1988; Sawilowsky, 2009). Thus the result is that on average, there was no significant difference in baseline assessment scores for participants in Group I versus participants in Group II; \( t (16) = -0.218, p = .830 \) at \( \alpha = .05 \) (please see Appendix H). This result confirmed expectations for the baseline assessment scores.

**Between Group Assessment Analysis Round One and Two**

The participants randomly assigned to Group I then performed the educational unit for four weeks (initial experimental group), while the participants assigned to Group II received the curriculum and course content as always delivered (initial control group), per the switching-replications experimental design and timeline as already described in the preceding sections. Round One of assessment testing was performed using the same tools as in baseline testing, after completion of the four-week educational unit by Group I. The independent sample \( t \)-test was used to test for a significant positive difference in
Round One assessment scores comparing the scores of participants in Group I (the initial experimental group) to the scores of participants assigned to Group II (the initial control group). Test hypothesis (one-tailed, or directional): Students in Group I scored significantly higher in Round One assessment scores than students in Group II (expected result). Null test hypothesis: There was no significant difference in Round One assessment scores between the two Groups.

For this independent sample $t$ – test, the computed $p$ value (Sig. 2-tailed) is .003; with a one-tailed hypothesis this $p$ value must be divided by two. The one-tailed $p$ value of .0015 is much less than the alpha value of .05, indicating significance, so the decision must be to reject the null test hypothesis. The mean Round One assessment score for the experimental Group I was 84.25, for the control Group II 50.56; when considering the direction of the means the finding is that there are higher scores for Group I compared to scores for Group II in Round One of assessment, supporting the test hypotheses and the decision to reject the null test hypothesis. The critical value of $t$ for $df = 15$ at $\alpha = .05$ is 1.753. With the computed value of $t$ at 3.469 much greater than the critical value of $t$ the decision to reject the null hypothesis is supported, that there was no significant difference in Round One scores between groups. The decision to reject the null hypothesis is also supported by the large effect size found. Computation of Cohen’s delta ($d$) for effect size for this independent sample $t$ – test resulting in a very large effect size of 1.68 ($> 1.20$), that is, the level of the independent variable (Group I or Group II) had a very large effect on the dependent variable (Round One assessment score). Thus the result is that on
average, participants in Group I scored significantly higher in Round One assessment scores than participants in Group II; \( t(15) = 3.469, p = .0015 \) at \( \alpha = .05 \) (please see Appendix H). This result matched expectations for the Round One assessment scores.

Per the switching replications experimental design, the participants in Group II then performed the educational unit for four weeks, while the participants in Group I became the control group but still received the curriculum and course content as always delivered, as already described in the preceding sections. Round Two of assessment testing was performed using the same tools as in baseline testing after Group II completed the four-week educational unit. The independent sample \( t \)-test was used to test for any significant difference in Round Two assessment scores between the students in Group I and those in Group II. Test hypothesis (two-tailed, non-directional): there was a significant difference in Round Two assessment scores between the students in Group I and those in Group II. Null test hypothesis: There was no significant difference in Round Two assessment scores between Groups (expected result).

The computed \( p \) value (Sig. 2-tailed) of .102 is much greater than the alpha value of .05 divided by two, equals .025 (for the non-directional test hypothesis). Therefore, the decision must be to not reject the null hypothesis (fail to reject). The critical value of \( t \) for \( df = 15 \) at \( \alpha = .025 \) is 2.131 (+, -). With the computed value of \( t \) at -1.742 within the range of the critical value of \( t \) at + or - 2.131 we must fail to reject the null hypothesis that there was no significant difference in baseline scores between groups. Computation of Cohen’s delta \( (d) \) for effect size for the independent sample \( t \)-test reveals a large
effect size of .846, e.g. there is a large effect on the dependent variable (the mean Round Two scores of the two groups) by the independent variable (group assignment), which is consistent with the observed but still statistically insignificant difference between the group means. Thus the result that on average, there was no significant difference in Round Two assessment scores for participants in Group I versus participants in Group II; $t(15) = -1.742, p = .102$ at $\alpha = .05$ (please see Appendix H). This result matched expectations for the Round Two assessment scores based upon the converge-diverge-reconverge pattern already described, with the difference in group means attributed to long-term degradation of learning effects in Group I.

An analysis of total time to complete all views was done for the Round Two assessment. Test hypothesis (two-tailed, non-directional): there was a significant difference in Round Two time to complete all views between the participants in Group I and those in Group II. Null test hypothesis: There was no significant difference in Round Two time to complete all views between Groups (expected result). The mean time to complete all views for Group I was 37.25, for Group two, 43.00. The computed $p$ value (Sig. 2-tailed) of .586 is much greater than the alpha value of .05 divided by two, equals .025 (for the non-directional test hypothesis). Therefore, the decision must be to not reject the null hypothesis (fail to reject). The critical value of $t$ for $df = 15$ at $\alpha = .025$ is 2.131 (+, -). With the computed value of $t$ at -.557 falling within the range of the critical value of $t$ at + or -2.120 we must fail to reject the null hypothesis that there was no significant difference in baseline scores between groups.
The decision to fail to reject the null hypothesis is supported by the small effect size found. Computation of Cohen’s delta \( (d) \) for effect size for the independent sample \( t \)–test shows a small effect size of .273, e.g. there is little effect on the dependent variable (the mean time to complete simulator views of the two groups) by the level of the independent variable (group assignment) and supports the decision to fail to reject the null hypothesis. Thus the result for the test for evidence of a significant difference in time to complete simulator views in Round Two of assessment between Group I and Group II is that on average, there was no significant difference in Round Two time to complete all simulator views for participants in Group I versus participants in Group II; \( t(15) = -.557, p = .586 \) at \( \alpha = .05 \) (please see Appendix H). This result matched expectations for the Round Two assessment time-to-complete metrics.

**Summary of Assessment Results Between Groups**

There was no significant difference in baseline testing scores between the two groups. For assessment Round One, the participants in Group I scored significantly higher in Round One assessment scores than participants in Group II; this was expected as Group I had just completed the educational unit. For Round Two of the assessments, there was no significant difference in Round Two scores for participants in Group I versus participants in Group II. This was expected, as now the Group II participants had completed the educational unit. There was no significant difference in Round Two time-
to-complete simulator views for participants in Group I compared to participants in Group II. This was expected, as Group II participants had completed the educational unit.

All of the above were expected results, per the converge-diverge-reconverge pattern seen in a switching-replications experimental design as described by Trochim and Donnelly (2008) and discussed previously. A graph of the mean scores by group and assessment round provides a visual representation of the converge-diverge-reconverge pattern, as seen in Figure 4 below.

![Line plot of mean scores by group versus assessment round.](image)
This line plot shows the expected pattern of diverge-converge-reconverge, representing a near-textbook instance of the initial experimental group initially exceeding the performance of the control group due to the effects of the experimental educational unit, then the initial control group catching up to the performance of the initial experimental group as they too, performed the experimental educational unit in the switching replications design.

The sole unexpected result was the drop in mean scores seen in Group I in Round Two of assessment. This was attributed to long-term degradation of learning effects of the use of the simulator, and would be easily addressed in an actual implementation.

**Within-group Assessment Analysis Results Round One**

Within-group analysis was performed to quantify performance of the participants within each group. The dependent sample t-test was used to compare performance for both groups to their own baseline performance after Round One of the experimental design and assessment. For Group I (the initial experimental group) the mean baseline assessment score was 10.00, with a mean Round One assessment score for Group I of 29.50. A very large effect size of 3.15 was computed. The paired-samples test results show a significant positive average difference between baseline and Round One assessment scores for Group I participants ($t_{7} = 8.914, p < 0.001$). On average, Round One scores were 19.50 points higher than baseline scores for Group I participants (95%
CI [14.32, 24.67]) (please see Appendix I). This was an expected result. The paired-samples test indicates a statistically significant positive and very strong effect of the treatment (the educational unit) on the dependent variable, the scores of the Group I members.

The dependent sample $t$-test was used for Group II (the initial control group) to quantify significant positive improvements in performance comparing baseline scores to assessment scores after Round One of the experiment. The mean baseline assessment score for Group II was 10.22, with a mean Round One assessment score for Group II of 17.66 and a moderately large effect size of 1.52, this paired-samples test indicates a statistically significant positive and moderately strong effect of the regularly-delivered course content on the dependent variable, the scores of the Group II members. There was a significant positive average difference between baseline and Round One assessment scores for Group II participants ($t_8 = 4.556, p = 0.002$). On average, Round One scores were 7.44 points higher than baseline scores for Group II participants (95% CI [3.67, 11.21]) (please see Appendix I). This was an expected result, and follows the converge-diverge-reconverge pattern discussed earlier. The expected result is a statistically positive improvement, just not as large of an improvement as seen by Group I, as the Group II participants still received the normally delivered content and learning opportunities of the standard course, while Group I had executed the educational unit including the use of the simulator.
Within-group Assessment Analysis Results Round Two

The dependent sample $t$–test was used for within-group comparison of assessment score performance after Round Two of the experimental design. Scores achieved by members of both groups in Round Two assessments were compared to their baseline scores, and to their Round One scores. There was a significant positive average difference between baseline and Round Two assessment scores for Group I participants ($t_7 = 10.333, p = 0.000$). On average, Round Two scores were 15.00 points higher than baseline scores for Group I participants (95% CI [11.56, 18.43]). A huge effect size was seen of $d = 3.65$ (please see Appendix J). This was an expected result.

There was a significant negative average difference between Round One and Round Two assessment scores for Group I participants ($t_7 = -2.496, p = 0.0205$). On average, Round Two scores were -4.50 points lower than Round One scores for Group I participants (95% CI [-8.76, -2.37]). A large effect size was seen of $d = 0.88$ (please see Appendix J). This was not an expected result, however, combined with the earlier independent $t$–test results of no significant differences between Groups in either scores or times in Round Two testing, this is deemed to be a result of minor concern but of interest nonetheless, that will be discussed further.

There was a significant positive average difference between baseline scores and Round Two assessment scores for Group II participants ($t_8 = 11.385, p = 0.000$). On average, Round Two scores were 19.11 points higher than baseline scores for Group II
participants (95% CI [15.24, 22.98]). A huge effect size was seen of $d = 3.79$. This was an expected result.

There was a significant positive average difference between Round One and Round Two assessment scores for Group II participants ($t_8 = 4.730, p = 0.001$). A very large effect size of $d = 1.58$ was seen. On average, Round Two scores were 11.67 points higher than Round One scores for Group II participants (95% CI [5.97, 17.35]). This was an expected result.

**Summary of Assessment Results Within Groups**

Statistically significant positive results were found for all within-group analyses, as expected, except one; the comparison of Round Two scores to Round One scores for Group I yielded a slight but still significant negative average decrease in scores of -4.50 points. This was not an expected result. After consideration, this result was attributed to the diminishing of the positive benefit of the use of the simulator in the educational unit over the four-week period that Group I acted as the secondary control group in the switching replications design. In other words, there was some degradation of performance due to long-term loss of the effects of the simulator use. However, the Round Two scores for Group I were still significantly much higher than their baseline scores, indicating a still significant and elevated positive cumulative educational effect. Additionally, consideration must be given that, as discussed in the previous section, the
independent sample $t$–tests for between-group comparison of Group I versus Group II scores and times in Round Two of assessment indicated no significant difference in either average scores or time-to-complete between participants in the two groups. Therefore, although there was degradation in performance within Group I’s scores, overall there was not a significant difference in performance by Group I in comparison to Group II participants.

True integration of the educational until into a course or series of courses could help mitigate these short term effect losses. By engaging the students in longer, continuing, or repeated sessions of increasing complexity, or in the study of additional and more-difficult views of the heart beyond the six basic views used, the loss of educational effect would likely be minimized. Even with the decrease in average scores in Round Two for Group I, the results all still followed the expected pattern of converge-diverge-reconverge, and reinforced the evidence of positive effect by the educational unit on learning outcomes. The results confirmed the positive effects of both the educational unit (the treatment) and the normally-delivered course content, both of which had positive effects on the dependent variables (the scores of the group members), with the educational unit embedding the use of the simulator having the greatest positive effects, as expected (please see Appendix J for details).
Additional Confirmation of Significance for Round One

As an additional statistical test to confirm the existence of statistically significant results, the ANCOVA analysis was performed for pretest/posttest scores for Group I and Group II. ANCOVA tests for a statistically significant difference in a dependent variable (post-test scores) between the levels of an independent variable (the experimental groups) after controlling for a covariate (the pre-test scores). ANCOVA was used to test for a statistically significant difference in Round One scores as the dependent variable, between the levels of the control group versus the experimental group (Group I versus Group II, the independent variable) after controlling for the pre-test (baseline) scores, the covariate. Test hypothesis statement (one-tailed, directional): After controlling for the pre-test scores (baseline) there was a statistically significant positive difference between Group I and Group II post-test (Round One) scores. Null test hypothesis: After controlling for pre-test there was no significant difference in post-test (Round One) scores based upon Group level. ANCOVA assumptions were satisfied for no significant difference between groups for pre-test (baseline) scores by performing an ANOVA test for the baseline scores of the two groups, and by testing for homogeneity of regression for between-subjects effects. Both assumptions were readily satisfied (please see Appendix K for details). The ANCOVA analysis was then run, with Levene’s Test of Equality of Error Variances showing that there was no significant difference in error between the two groups. The computed $p$ value for between-subjects effects for Group is
.001 which is statistically significant at the value of .05 so the decision must be to reject the null test hypothesis. The Partial Eta Squared value for Group of .575 indicates that variance in membership from one Group to the other accounts for 57.5% of the variance in the dependent variable, the post-test (Round One) assessment scores. This supports the decision to reject the null test hypothesis. Considering the direction of the means with mean values of 84.606 for Group I and 50.239 for Group II, there are higher post-test scores for Group I compared to the scores for Group II in Round One of assessment. This supports the directional test hypotheses and the decision to reject the null test hypothesis.

Findings Statement for the ANCOVA test: When controlling for the pre-test scores (baseline assessment scores) there is a statistically significant positive difference in post-test scores (Round One) with higher Group I scores versus the scores in Group II (please see Appendix K). This result supports and confirms the results of the between-groups analysis using the independent sample t-test as described earlier.

Correlational Analyses

A large number of bivariate correlational analysis tests were performed to explore the possible predictive/criterion relationships between simulator performance and clinical performance of the same skillset on an actual person. To begin, an expected negative correlation was found between total time to complete all views using the simulator, and the overall score for all participants in Round Two assessments, with a Pearson $r = -.183$
and weak negative correlation as seen in the scatterplot in Figure 5 below. This portrays a negative correlation that, as the time to complete goes down the assessment score tends to go up, an expected result.

![Figure 5: Scatterplot of time to complete simulator views vs. Round Two scores.](image)

Of additional interest were the large number of weak to mild positive correlations that were seen for many simulator-based criteria for performance (raw or combined scores, time to complete, per view time) and the corresponding criteria from clinical assessment or the clinical practicum. The best of these numerous weak positive correlations is represented in the scatterplot shown in Figure 6 below, showing a mild
positive correlation between the combined score on the simulator to the clinical practicum total score, with a Pearson $r = .185$.

![Figure 6: Scatterplot of simulator combined score vs. clinical practicum score.](image)

Since adding together three assessment criteria into a combined score netted the most-promising correlation, an additional method was used in an effort to find further positive correlations. A principal component factor analysis extraction method was performed using average time to complete, total time to complete, and Round Two scores as components (please see Appendix L). The extraction yields a single determinant factor for use as a predictor of criterion outcome preserves the degrees of freedom versus
using the various original components individually. The result was designated the Simulator Performance factor. A mild positive correlation with Pearson $r = .227$ between the Simulator Performance Factor and the clinical practicum score is found, shown in Figure 7 below.

![Scatterplot of Simulator Performance factor vs. clinical practicum score.](image)

**Figure 7: Scatterplot of Simulator Performance factor vs. clinical practicum score.**

However, even using the Simulator Performance factor, it was still difficult to find any further meaningful positive correlation to clinical performance; most analyses portrayed weak if any positive correlation. As a final effort another principal component factor analysis extraction method was done to combine data from the various clinical performance measurements into one, resulting in the Clinical Performance Factor (please
see Appendix L). Using the two extracted factors, a small positive correlation was found between the Simulator Performance Factor and clinical performance as measured by the Clinical Performance Factor, with a Pearson $r = .132$ and shown in Figure 8 below.

![Figure 8: Scatterplot of Simulator Performance factor vs. Clinical Performance factor.](image)

Other, unexpected negative correlations were found. For example, a weak negative correlation was seen between time to complete all views on the simulator and the time to complete the same views in a clinical setting, with a Pearson $r = -.022$ (please see Appendix L for details). This seemed counterintuitive and was unexpected, but will be discussed further in the next chapter.
Summary of Correlational Analysis

A large number of weak to mild positive correlations were found when looking at simulator-based performance as a predictor of clinical performance. Some unexpected negative correlations were found as well. The inability to find strong positive correlations between the various simulator-based performance measures and the corresponding activities in a clinical setting was at first puzzling, and was contrary to initial expectations. However, after much consideration this was seen to make sense due to important characteristics and limitations of the simulator itself that will be discussed in detail in the following chapter in the Conclusions section.
Chapter 5

Conclusions, Implications, Recommendations, and Summary

The overall research goal to determine if the time, expense, and effort needed to formally integrate the simulator into an existing course was warranted was accomplished and affirmed with overall statistically significant positive findings indicating that formal integration is worth the effort. There were strongly statistically significant affirming results in many areas, and a few cautionary indications in others, but the overall outcome of the study was positive. A discussion of the pluses and minuses of the experimental design, the educational unit, and the simulator features that contributed both directly and indirectly to the results follows.

Conclusions

Answering the research questions. The research questions were answered by the results of the following activities. For the first research question, what is the current state of the literature regarding the integration of simulation into existing curricula, including best-practices and identified gaps in the research regarding such implementations, the literature review resulted in a positive answer to this research
question in the form of the information desired. The goal of the literature review was met, with an exploration and discussion of the published efforts of others on the integration of simulation technology into existing curricula through the application of established learning theory and ID theory. Specific literature identified focused on the use of applied constructivist learning theory to develop such instructional units, and on integration of simulator technology into healthcare related curricula.

The second research question, what are the foundations of the proposed educational unit from both a learning theory and ID theory standpoint, and the applicable ID methods to be used for the research, was answered positively by the discovery and discussion of the selected aspects of applied constructivist learning theory, and of the found instances of ID selected for further use. The selected ID models of Jonassen (1999) and Colón et al., (2000) were used in a combined ID template for the experimental educational unit, blending the excellent constructivist features of both into a new and powerful construct, a most emphatic positive answer to this research question.

The third research question, what are the resulting effects on learning outcomes that can be attributed to the educational unit, as analyzed by quantitative methods, was answered by the use of known parametric quantitative statistical analysis methods to compare standardized learning outcomes for the effects of the educational unit. Both between-group and within-group analyses of performance measurements were used, comparing baseline performances to post-educational unit performance data for both groups in the switching-replications experimental design. These analyses answered the
research question with positive results affirming the beneficial effects of the simulator use within the experimental educational unit.

And finally, the fourth question, what evidence is found supporting the existence of knowledge transfer from simulator skills to clinical skills was addressed by the correlative analysis of post-test results for both groups to the performance of the same participants in a clinical setting where the same procedures as practiced on the simulator were performed in a clinical setting on actual persons. This question was answered positively, but only in a mildly affirmative way, as only a small positive correlation was seen between simulator-based outcomes and the performance of clinical skills.

The experimental design. The use of the switching-replications experimental design was a complete success. At the initial presentation of the study and request for volunteers, the unique structure of the experimental design to be used was explained in detail. Quite a few students asked further questions regarding the design, and all seemed happy at the thought that everyone would get a chance to use the simulator during the study, regardless of whether they were in the initial control group or initial experimental group, as in the switching-replications design the two groups switch roles after the first round of the experiment. It is thought that careful and thorough explanation of the switching-replications experimental design appealed to this group of learners’ sense of fairness and thus strongly compensated for the very strong possible social threats to internal validity discussed previously. There was still some mild competition evident, but it always seemed to be of a healthy sort and was nowhere near the vindictive tone and
aggressive attitude that was seen during the pilot run-through in the Spring. The appeal of this experimental design to the concept of “fairness” that seems to permeate the thought-processes of this generation of learners is perhaps of critical importance to other researchers and should not be ignored especially in an educational setting. The minimal extra effort to repeat the experimental phase with the second round of participants was well worth the trouble, as it also served to effectively double the dataset produced for analysis; researchers faced with small potential populations from which to draw participants should take note of and consider using this design.

**The educational unit.** The use of the simulator was embedded in the constructivist-guided features of the educational unit as supporting structure for learning. Some of the features of the educational unit were valued more than others by the participants, and for varying reasons. One of the most valued constructivist features of the educational unit was the scaffolding support process that was used to provide assistance to the participants in their use of the simulator. The initial heavy support presence needed by many students the first time they used the simulator quickly gave way to only an occasional question, usually within the first two or three sessions. Basic how-to type questions quickly gave way to inquiries about details of the anatomical structures of the heart, the interaction of various parts of the heart anatomy, or additional features of the simulator. In both groups, soon the usual basic simulator operation and ultrasound imaging questions were quickly supplanted by questions leading to other areas that often proceeded to the point of eventually needing to redirect the participant back to
the task at hand. However, this kind of interaction was valued, as it showed true engagement and curiosity by the participants and was never actively discouraged. Based upon comments by participants, the freedom of inquiry to explore using the simulator technology freely and openly yet within the supportive structure of the overall educational unit was of great value to the participants as learners.

Less well-received constructivist features of the educational unit were the portfolio, Web quest, and learner reflective blog. These were almost universally ignored in the experimental phase by each group, an outcome foreshadowed and forecast by the pilot run-through earlier in the Spring. As in the pilot, the nearly universal response from participants as to why these activities were not used was that they were perceived as extra work and were not valued unless they were going to be graded. These features, especially the portfolio and Web quest, could easily be assessed for a grade if the educational unit were embedded in a course. In fact, these constructivist ID elements were designed with assessment in mind, if fully implemented in an actual course. However, as described earlier and to avoid any difficulties of undue influence to a potentially vulnerable population, it was communicated from the beginning that nothing the participants did or did not do would affect their grade in the actual course. This was repeated to all students starting with the initial recruitment of volunteer participants and distribution of the IRB consent form, and continuing throughout the duration of the experiment. In an attempt to counteract the possible effect of a no grade = no worries attitude toward these activities, the importance of performing these activities was
continually mentioned to all participants during all interaction. Friendly reminders were posted in the simulator room and emails sent to the group participants during their respective turns in the experimental unit. However, at least for these participants in this setting, the effect of not being evaluated for a grade was seemingly more powerful even than their volunteerism. This is an interesting outcome that would be worth exploring in a future, related study. It is also a cautionary note to others who may wish to implement what may be thought of as worthwhile but possibly time-consuming ID features into their own studies, as this effect seemed to overpower even the otherwise excellent participatory attitudes of the participants. The time-spent perception combined with the no grade effect, seemed to be a powerful set of influences against the performance of some activities.

Features of the educational unit that received modest support were the weekly image-based quiz and the discussion board. Most participants found the oral, image-based weekly quizzes useful low-stakes measurements of their own progress. Since these assessments were not graded they apparently did not have the burden of anxiety that often accompanies most assessments, with several participants commenting that they were even fun. In contrast to the preceding discussion of the no-grade effect, this illustrated the fine line between the need to assess learners and the ability to provide effective learning in order to achieve desired educational effects.

A few students used the discussion board to provide feedback and suggestions throughout the experiment. Although participation was low overall, thoughtful
comments and suggestion were made by some participants and in some cases acted upon. However, the anticipated asynchronous dialogue of sharing tips or helpful hints between learners as they moved through the educational unit never occurred. This could be an effect of the limited duration of only four weeks of each phase of the experiment. Perhaps an experimental phase of longer duration, or one that did not have the same level of strong scaffolding support for the learners, would have generated more discussion traffic of the anticipated nature, as the discussions then would have filled a need for helpful information.

Another takeaway for the ID structure and features of the educational unit was that the scaffolding support required by the learners was not nearly as complex in content-level, nor as time-intense, and did not need to remain at a high level of support for as long as originally planned. This could have been a result of the advance preparation in the form of the basic training on the simulator that was held for all students prior to the experiment. The additional support available in the form of quick-reference sheets, online tutorials, and the availability context-specific help within the simulator itself may have also reduced the learners’ need for intensive and extensive scaffolding support.

One of the biggest factors contributing to the success of the educational unit was that the participants were excited, engaged, and eager to learn on the simulator. This type of human factor may be hard to achieve and influence in all cases. However, much effort was expended to foster excitement and interest, including the initial briefing over pizza
lunch, the award at the end of gift cards, continual positive reinforcement during scaffolding sessions, and so on. All of the preceding discussion points are based upon comments made by participants during the experimental phase of the study and anecdotal observations gathered during informal Q&A sessions at the end of each four-week phase of the experiment. Such participant feedback would have been better captured via more formal qualitative methods such as structured interviews, think-aloud sessions, or satisfaction surveys. Such methods were however beyond the current scope, but may be undertaken in the future as an extension of this work or would be an excellent topic for future related research.

**The simulator.** The simulator was an excellent tool for participants. With it, they learned the anatomical structures of the heart, and to recognize, identify, and attain both anatomical and ultrasound views of the heart. The anatomical representation of the heart is an overwhelming strength of this particular simulator. The detail and anatomical accuracy of the 3D model of the human heart that is the core of this simulator is without equal. Please see Figure 9 below, for a simulated view of a particular image approach to the heart, and Figure 10 for the same view on an actual person via an ultrasound machine.
Figure 9: Simulated ultrasound view of the human heart.

As can be seen above the simulator does an excellent job of portraying the anatomical structure and views of the human heart. Unfortunately, the simulator fails to simulate an actual ultrasound machine as applicable to the user interface (UI). It is
believed that the unexpected results seen in certain areas of the quantitative analysis stem from this fundamental UI design shortcoming. To illustrate this point, recall from the Results chapter that the participants’ performances on the simulator were at best only a mild predictor of their performance when using actual medical ultrasound imaging equipment on a live person to obtain the same imaging result. There were even some unexpected negative correlations found. These findings were contrary to initial expectations, but after further consideration made sense due to the factors discussed in the following Implications section.

**Implications**

Implications for the use of the switching replications experimental design are that care should be taken to clearly communicate to all potential participants. Such explanations seem to appeal to participants’ sense of fairness and are a compensating factor for possible social threats to internal validity, that can be very strong and have detrimental effects. A researcher in an educational setting considering a randomized trial involving a control and experimental group would be well-served to at least consider the possibility of using a switching replications design.

It is possible that the user interface (UI) design of the simulator used was both the direct and indirect source of some of the unexpected results seen. Direct effects on the results may have occurred as a result of the way the simulator trains the user to adjust the
image by methods and controls that are unique and proprietary to the simulator and only loosely based in the reality of an actual ultrasound imaging machine. Thus when later faced with an actual ultrasound imaging machine, the controls to manipulate basic image functions such as depth, brightness, etc. are unfamiliar to the learner and cause hesitation. A transition period from the controls and methods used on the simulator to those of the actual imaging equipment is required. A visual comparison of the UI physical controls of an actual ultrasound machine as used by the learners in the clinical lab setting (Appendix M) with the physical and screen-based UI controls of the simulator (Appendix N) illustrates the problem. The respective UI controls on each machine are superficially similar in appearance due to the presence of a standard QWERTY computer keyboard on each device, and a simulated transducer on the simulator that very closely matches the appearance of the actual ones on the ultrasound machine. However, the main controls of the ultrasound machine are only incompletely replicated by on-screen icons on the simulator, and accompanied by many proprietary icons of non-intuitive appearance and function. The third image provided in Appendix N is a close-up of some of these on-screen icons, that must be activated with a mouse-click (please see Appendix N). There is little shared between these icons and their counterpart, if it even exists, on an actual ultrasound machine, including the manner that they are activated. There are typically no mouse-clicks involved on an ultrasound machine at the basic control level, instead there are large, easily manipulated buttons to push and knobs to turn. It may be that these fundamental differences in the UI lead to hesitation and confusion, especially for a
beginning-level learner. Thus the very real and powerful advantages of learning about heart anatomy and ultrasound views via the simulator are possibly counteracted by the complexity and confusion of the UI. The learner must then overcome these UI differences when transitioning from the simulator to an actual imaging ultrasound machine in a clinical setting. This could be another area for further future research, to determine the effects of a more-realistic user interface on some of the learning outcomes that could be affected by the shortcomings in UI design of the simulator.

Indirect influence was possibly indicated in that, once using an actual ultrasound imaging machine on a real person, many of the learners seemed to struggle for inordinate amounts of time. When questioned, the reply was often that they were trying to get the same level of perfection in the image on a real person that they were used to seeing when using the simulator. A high level of image clarity is often simply not possible on a real person. On the simulator, they had been inadvertently trained to expect perfection and when perfection was not available when examining an actual person, they became frustrated and tried even harder to get it since that was what they were used to seeing. Granted that these participants were all beginning learners in this particular field, and that a more-experienced learner would likely not make this type of error in what is essentially a time management issue. A more-experienced learner would likely recognize the need to balance between time efficiency and image quality. However, this seemed to be a possible effect of learning on the simulator that could explain the large time differences between performance on the simulator and clinical performance, and one that could be at
least partially overcome with further additional simulator design features to be discussed in Recommendations.

A third area of the simulator that was thought to contribute to the observed performance differences concerns two reality or fidelity characteristics that involve the difficulty, or this case lack of difficulty, in image acquisition. As seen previously in Figures 9 and 10, the simulator does an excellent job of portraying a simulated ultrasound view of the human heart. A tactiley-realistic mannequin with flesh-like chest and abdomen and underlying bony-feeling anatomy, combined with corresponding simulated image effects such as shadowing when passing over the ribs in simulated ultrasound mode, does add to the reality of the simulation. However, there are two very important fidelity features missing. Most important of these is that the simulator does not breathe, either in a physical sense or via virtual representation. A real person not only has a beating heart, that the simulator does portray very nicely in a normal sinus rhythm of approximately 60 beats per minute, but also breathes more or less regularly, fifteen to twenty times per minute at rest. The attendant motion of breathing causes the ribcage, chest, abdomen, and the heart itself to rise and fall. This motion is a very large part of the difficulty in obtaining correct views of the heart when examining an actual person. In contrast, the simulator and the simulated image remains perfectly still except for the regular rhythmic contractions of the beating heart model. Asking the patient to temporarily pause their breathing on an inhalation, exhalation, or with a partial breath-
hold is a common and valuable technique used during an actual ultrasound examination, for this very reason.

Similarly, varying levels of viewing difficulty exist from person to person when actually examining them via ultrasound. Some people are simply friendlier than others to the wavelengths of sound used in medical imagery. This occurs due to multiple reasons beyond the scope of this discussion involving the physics of the propagation, reflection, attenuation, scattering, and other behaviors of sound moving through various bodily tissues, a medium of varying composition.

Both of these reality concerns could be incorporated in the simulator, perhaps as separate features that could be dialed-in with increasing levels of more life-like action and increasing attendant difficulty. This would allow a raw beginner to practice with a beating but not otherwise not moving and crystal-clear heart. As learner proficiency progressed, difficulty could be increased in the form of adding and increasing breathing movement and/or image clarity issues to prepare the learner more effectively for performing imaging tasks on an actual person.

The extraordinary difference in the time used to obtain the same images in a clinical setting in comparison to their time-to-complete on the simulator is difficult to explain other than by the possible effects outlined above. Future follow-on research could include qualitative investigation of the learners’ thoughts and perceptions as they use actual ultrasound machines after practicing on the simulator, perhaps through think-aloud and active interview techniques.
Recommendations

Recommendations to create a UI that more closely replicates the controls and functions of an actual ultrasound machine have been previously provided to the manufacturer of the simulator. The shortcomings of the existing UI design have been discussed with a team from the manufacturer during two sessions over the last two years, where their representatives attended courses and lab sessions, observed students using the simulator, and sought input on possible new developments in roundtable discussions. These efforts to collect product improvements were admirable, however they have so far yielded only software-based refinements. As of the latest product release the user still interacts with the simulator software via the proprietary and non-intuitive screen symbols, a small selection of virtual buttons on the screen and numerous menu pull-downs activated by mouse clicks, and keyboard commands, all of which have only a passing if any resemblance to the controls of an actual ultrasound imaging machine. Improvement to the design of the UI to more closely represent the controls of an actual ultrasound imaging machine would be the most important recommendation for the creators of the simulator. This would be followed closely by increasing imaging reality levels by incorporating the ability to add and adjust the levels of breathing movements and imaging difficulty. These are the key improvements recommended for the simulator itself, which will be passed on to the manufacturer.
Recommendations for fellow educators who may wish to use the educational ID unit for their own incorporation of simulation technology into their courses, include that any such implementation would be well-served to choose the constructivist features that can be actively assessed and incorporated into the grading structure of the host course. Such assessment should be organized around a solid core of scaffolding support for their students. The scaffolding methodology provided a framework, just like its real-world namesake, that allows for the building of knowledge, the exploration of concepts, and the engagement of learners. Complementary constructivist activities for social and reflective learning such as blogs, discussions, discovery learning activities, etc. should be used, but used with a realistic outlook, in that if they are not actively assessed, the learners involved may deem them not worth the effort and time required to perform.

Recommendations for fellow researchers, especially in educational research, are to strongly consider the use of the switching replications experimental design. This is especially so if they anticipate likely positive results of the planned educational intervention, if their potential participant population from which volunteer participants are drawn is socially connected, and if there is a need for maximization of data for analysis. With both randomized groups eventually performing the intervention in this type of experimental design, outcome data is literally double that of a simpler randomized controlled design, leading to more generalizable and useful results.
Summary

**The problem.** The problem identified was the attainment of meaningful integration of simulator technology into existing curricula. The lack of meaningful integration is a commonly-seen problem with the use of simulator technology in health care education (McGaghie, Issenberg, Petrusa, & Scalese, 2010; Wittels, Takayesu, & Nadel, 2012; Motola, Devine, Chung, Sullivan, & Issenberg, 2013). Many educational institutions that have acquired simulation technology have difficulties attaining meaningful improvements in learning outcomes, transfer of learning from the classroom to the clinical setting, and full utilization of the technology (Arthur, Levett-Jones, & Kable, 2013; Cook et al., 2013). A tendency has been identified for simulator technology to be treated as an add-on to the curriculum with little effort to fully integrate its use into the learning environment (Masters, 2014). The generally-held positive outlook on the effectiveness of simulation use in health care education is not well-supported by rigorous empirical research, specifically into areas of effective ID (McGaghie et al., 2010).

Health professions educators and administrators have few or no guidelines for the design of ID systems to effectively utilizing simulator resources that are usually capital-intensive and require evidence of a reasonable return-on-investment. In addition, the majority of research on the use of simulation technology in health care education involves medical school settings for the training of physicians, with little available in the literature.
involving the education and training of allied-health practitioners (McGaghie et al., 2010).

**The research goal.** The research goal was to add to the body of knowledge of ID, a working template for the successful integration of advanced simulation technology into an existing health care education curriculum. The specific curriculum in this case was for cardiac sonography, with the ID using applied constructivist educational techniques. This addition to the body of knowledge will benefit both the theory and practice of ID, as future researchers in ID theory may find this example of a practical application of a working integration to be instructive, and practitioners may find the ID template to be a useful model, adaptable for their own circumstances. Similarly, educators in health care programs who wish to integrate simulations technology into their own curricula may benefit from use of the ID template for their own integration efforts of simulation technology.

 Statistical analysis of learning outcomes and correlation to clinical outcomes provided supporting evidence of the guiding research hypothesis: students who perform the experimental educational unit will achieve statistically significant higher scores on standardized assessments. If the findings did not support the research hypothesis, then the null hypothesis of: there was no statistically significant difference in standardized assessment scores between students who have received the experimental intervention and those who have not, would be supported.
Stated in an even more fundamental way, it was sought to determine if the formal integration of the use of the simulator technology into the existing course curricula via the constructivist-based instructional unit was worth the time, effort, and expense required. This was determined by the findings of statistically significant improvements in educational outcomes and a positive correlation to clinical outcomes, indicating knowledge transfer from use of the simulator to the real world.

The research questions. The following research questions were addressed and answered:

1. What is the current state of the literature regarding the integration of simulation into existing curricula, including best-practices and identified gaps in the research regarding such implementations?

2. What are the foundations of the proposed experimental educational unit from both a learning theory and ID theory standpoint, and the applicable ID methods to be used to design and develop the instructional unit?

3. What are the resulting effects on learning outcomes that can be attributed to the experimental unit, as analyzed by quantitative methods?

4. What evidence is found supporting the existence of knowledge transfer from simulator skills to clinical skills?
The state of the literature involved in research question one and the foundational design issues of research question two were both positively answered. The literature review identified the published efforts of others on the integration of simulation technology into existing curricula through the application of established learning theory and ID theory. Specific literature identified focused on the use of applied constructivist learning theory to develop instructional units, and on integration of simulator technology into healthcare related curricula. Concurrent to the literature review, the discovery and discussion of selected aspects of applied constructivist learning theory, and of the found instances of ID selected for further use as a combined ID template for the experimental educational unit took place. The blending of the constructivist features of both of the ID models of Jonassen (1999) and Colón et al., (2000) into a new and powerful construct, resulted in a most emphatic positive answer to research question two.

Positive findings of statistically significant improvements in educational outcomes answered research question three very strongly in the affirmative. The mild positive correlation found for simulator performance to clinical performance supported research question four affirmatively, though only weakly. All of these findings will be discussed in further detail in a later section of this Summary.

**ID of the experimental educational unit.** The experimental educational unit ID used applied constructivist techniques including scaffolding and cognitive apprenticeship in the integration of a high-fidelity human heart and TTE/TEE simulator into the existing
The simulator, known as HeartWorks (http://www.heartworks.me.uk), comes with rudimentary learning features that were incorporated into the experimental ID to develop a self-guided learning program that formally integrated the use of the simulator into the existing curriculum. The simulator was employed specifically to teach the basics of human heart anatomy and its proper imaging via medical ultrasound. The educational unit was launched only after all participants had first completed basic training on the use of the simulator itself. This minimized possible difficulties due to learner unfamiliarity with the system, addressing the issues brought forth by Ramdass (2012).

The experimental educational unit was a combination of two constructivist-based ID development models. After a review of the many published constructivist-based ID models, Jonassen’s (1999) PC³ development model, and the R2D2 development model of Colón et al., (2000) were selected. Instances were found of both models being used by others in learning situations, contexts, and environments similar to those extant in this investigation with resulting beneficial effects on learning outcomes and are thus considered to be proven instructional designs. Both models were used to develop the educational unit, as both bring different yet desirable and complementary constructivist design aspects to address the ID development needs.

The PC³ ID model was created by Jonassen in an effort to promote the practical use of constructivist learning theory in technology-aided learning environments in a manner meaningful to instructional designers (Jonassen, 1999). Jonassen was an early proponent of the adaptation of constructivist theory into ID, advocating the use of
appropriate foundational theories within the ID context to facilitate the best learning experience possible (Jonassen, 1999). Jonassen’s PC³ model uses scaffolding concepts to guide students in the interpretation and resolution of learning problems. The Problem (P) or question is outlined, with student understanding ensured by supporting background information. Then, technology based Cognitive tools (C¹) help the learner engage and interpret the Problem to enable the constructive learning experience. Such tools may include visualization and performance tools (e.g. simulations, virtual reality, modeling), knowledge modeling tools (e.g. mind- or concept-mapping), and information gathering or data mining tools. Additionally, Conversation (C²) and Collaboration (C³) tools are utilized to allow for the important social-learning aspects of constructivist theory, enable the co-construction of meaningful learning experiences related to the Problem.

The conceptual framework provide by the use of the PC³ model was further enhanced via the use of the R2D2 ID model of Colón, Taylor, and Willis (2000). The R2D2 model is itself based upon the foundational work of Willis (1995) who created an early ID model incorporating recursion and reflection as key components to a continual, iterative design process.

Willis asserted that there are three primary or first order principles of constructivist-based ID models: recursion, reflection, and participation (Willis, 2000). These features have been expanded and elaborated upon in the later R2D2 model. There are four underlying key themes within the R2D2 model: Recursion, Reflection, a non-linear iterative Development pathway, and the use of participatory Design. The R2D2
model allows the designer to continuously update and revise the instructional model based on feedback from the participatory group, and from experiences gained from use.

Combining Jonassen’s (1999) PC³ model and the Colón et al., (2000) R2D2 model resulted in a new unique ID framework, that was used to develop a Comprehensive Lesson Plan for a four-week long educational unit integrating the use of the simulator (Yoders, 2014). Subsequently, a condensed, Focused Lesson Plan was created, distilling participant time requirements and content delivery to a minimal three-week long core, but still retaining all of the important constructivist learning features.

After an ID affirming pilot run-through with the three-week Plan, described in more detail in a later section, additional adjustments were made to the experimental educational unit. The now revised Focused Plan went back to a four-week duration, as the consensus of the pilot volunteers was that three weeks may not be enough time to work with the simulator. Additional changes were also incorporated in the ways that participants used the simulator, in the logistics of participant sign-up for simulator sessions, the provisions for scaffolding support of the participants, and content added to the supporting materials.

Experimental design used. Next, the design of the experiment was considered and after much deliberation, a switching replications experimental design was used (Trochim & Donnelly, 2008, p. 234) with random assignment of participants to a control group and an experimental group, and three waves of assessments. In the switching replications design, the randomly assigned control and experimental groups both receive
an initial pre-test, then participate in the treatment or not depending on the group, followed by a first wave of post-testing. Then a replication of the treatment occurs with what was previously the control group, who after the first experimental phase switch roles and become another experimental group, with the previous experimental group now inactive. A final post-test is performed, including in this case a secondary clinical assessment.

The switching replications experimental design was chosen since students are the source of the volunteer participants, and there was a concern that those assigned to a control group may be denied possible positive benefits of the treatment. Pesiridis et al. (2015) discuss that when a treatment is desirable or likely to be beneficial, and when performed in a setting that may not be fully amenable to a no-treatment group, random assignment of participants to an early treatment group and one that initially serves as a control group, but will receive the treatment later may be desirable. Thus the switching replications design overcomes a major problem of the typical randomized controlled trial, since the need to deny the treatment (in this case, desirable training) to the control group is eliminated.

Additionally, there can be very real internal threats to validity that may occur due to the perceptions and resulting behaviors of participants assigned to a control group who are not isolated from the experimental group, but are socially connected (Gay, et al. p. 261). These threats may arise from the perception that those receiving the treatment in an experimental group are being afforded some disparate or unfair advantage over those in
the control group. As a result, behaviors in the control group such as compensatory rivalry, resentful demoralization, diffusion, or imitation of treatment, may result (Trochim & Donnelly, 2008, p. 171). By ensuring that all participants receive the possible benefits of the treatment, the switching replications experimental design both alleviates investigator’s ethical concerns over fairness or benefit issues and effectively controls for possibly very powerful social threats to internal validity.

As an added practical benefit, since all participants receive the treatment and are identically assessed, the switching replications design provides additional data for analysis. Such data can then be examined in at least two ways; comparisons of the pre- and post-treatment test scores of Group B and Group A both within groups and between groups; and the combined post-treatment scores of both groups in a correlational analysis for evidence of knowledge transfer of simulator skills to clinical skills.

**Execution: pilot run-through.** In the Winter term of 2016, a proof-of-concept or pilot study was performed to fine-tune the educational unit procedures and assessment instruments for the future, larger study targeted to involve the incoming class of May of 2016 and occur in the Summer term of 2016. A minimal experimental plan was used, with no pre- and post-test assessment. With a single group of nine (and later, ten) volunteers as the experimental group, the switching replications design of the full study was not utilized for the pilot activities. A tenth student volunteered in the first week of the pilot activities, after she stopped in to see what the volunteer students were doing on the simulator, and decided she wanted to participate after all.
Despite being of limited scope, all other planned activities were performed during the pilot, starting the week of March 27th, 2016, and for the following three weeks, culminating with a hands-on clinical assessment in late April 2016 as described below. Key aspects of the planned educational unit were still employed including scaffolding of learner support for participants, assessment via interactive image-based quizzes, online discussion board and blog for social interaction, social learning, and self-reflection, and the Web quest to allow for possible extended learning opportunities. Despite the limited number of participants and the limited nature of assessment during the pilot, valuable insights were obtained by the pilot run-through. As a result of the pilot, the assessment tools were fine-tuned, including additions to allow notation of the exam room used, time started and ended and elapsed time overall, and other minor but useful details. The run-through suggested areas for procedural and logistical improvements, particularly in the way learner support via scaffolding was handled, and certain time management and logistical issues.

Overall the ten volunteers did well on their clinical scores, and in comparison to the abstaining students, showed statistically significant positive differences in some aspects of their performance, although there were no differences in time performance criteria. This was perhaps due to confounding effects of social threats to internal validity, specifically compensatory rivalry. In this case the non-participants became competitive with the 10-participants volunteer group who were perceived as receiving special treatment. The non-volunteers compensated by trying to out-perform the volunteers
during testing. This threat to internal validity is controlled-for by the switching replications design (Trochim & Donnelly, 2008, p. 234) as used in the full experimental study. However, as already described the switching replications was not used for the pilot run-through.

**Lessons learned and preparations.** Preparation for execution of the full experimental unit during the Summer 2016 term took place in April and May of 2016, to involve the students of the incoming class of May 2016 (N = 19). As a result of the pilot, fine-tuning of procedures and scaffolding arrangements took place as previously discussed, and the assessment tools were reverted to the original six basic views, taking them back to a level suitable for beginners.

On Friday May 20th, 2016 an announcement and description of the study and a request for volunteers was made to the students. Eventually every student except one volunteered, resulting in (n = 18) participants. The sole hold-out said that she was concerned with the extra time commitment needed and chose to err on the side of caution by not participating.

On Friday May 27th, the 18 participants were randomly assigned to the initial experimental group (Group I) or to the initial control group (Group II) via a random sequence generator (please see Appendix F). Descriptive analysis of demographics comparing factors between the two randomized groups revealed a similar distribution of age range, averaging 23-28 years old for both Groups; and for education level, with a
factored analysis yielding a 1.44 average for Group I and 1.22 average for Group II, within a range of 1 equaling bachelor’s level and 2 equaling master’s level.

Execution: the full experimental unit. Baseline testing for knowledge of the six basic views of the heart and the ability to identify structural anatomy of the heart in these views was held on Friday June 3rd, 2016 for all 18 participants in both Groups. Scoring data was collected for later analysis, then the Group I students began the first of four concurrent weeks of simulator use per the experimental educational unit starting the week of Monday, June 6 through Saturday, June 11th, 2016. Group I students signed up for two one-hour-long or four 30-minute long simulator sessions per week. The fourth and final week of simulator use for Group I participants was Monday June 7th through Saturday July 2nd, 2016. The first round of testing was completed on Friday July 1st, 2016 for both Group I and Group II participants, using the same assessment tools as the baseline test.

At this point the Group I students had experienced both the benefit of the normally-delivered educational curricula plus the simulator use within the educational plan, while Group II student participants had only received the regularly delivered materials and methods. The expectation was for some positive improvements to be seen in Group II scores over their baseline scores, since they were still being taught via normal methods, but that the Group I scores would be significantly positively higher than both Group II scores and their own baseline performances. This is exactly the pattern seen after later data analysis.
Starting the week of July 5th through the 8th, the participants in Group II then performed the experimental educational plan and used the simulator for the following four weeks, through the week ending August 30th, 2016. During this four-week period, Group I participants did not use the simulator but did continue to receive all normally delivered instruction via regularly delivered materials and methods. This was per the switching replications experimental design previously described.

The final round of assessment testing for both Group I and Group II occurred on August 2nd, 2016. The same assessment tools used in both the June 3rd baseline testing and in the Round One testing on July 1st, were used in the testing on August 2nd. This was followed by hands-on clinical assessment on Friday, August 5th to assess knowledge transfer from the simulator skillset to a clinical performance. An additional source of data for correlational analysis and possible identification of long-term learning effects was the midterm exam for all students during the Fall term, held on October 27th and 28th, 2016. This midterm hands-on exam assessed the students’ skills in acquiring the same six basic views as those targeted during the experimental phase, thus providing another point of performance comparison.

All data was collected for later analysis, and after each session, an informal focus-group discussion was held seeking comments and feedback from participants. All participants were then given their choice of a $20 gift card from various establishments (Starbucks, Subway, Cold Stone Ice Cream, or movie tickets) that were paid for out-of-pocket as a personal thank-you and delivered to them by the last day of class for the
Summer term, Friday August 12th, 2016. Data organization and statistical analysis began the week of August 15th, 2016. Data analysis work was completed in late January 2017.

**Results.** Statistically significant positive improvements were seen in the educational outcomes of participants when measuring both the knowledge and application of heart anatomical structures and the various views used during examination of the heart with ultrasound. A strong finding is that the simulator used is exceptionally useful in the teaching and learning of anatomy and anatomically-related views and concepts, excelling as a simulation platform of the human heart.

Only mild positive correlations were found between performance on the simulator compared to performance in an actual clinical setting, with only a weak predictive value between the two. It was determined that for the predictive value to improve, the simulator is in need of changes to provide a more-realistic portrayal of the limitations of actual ultrasound imaging. Most critically, the simulator as used lacks a realistic user interface as a procedural simulator of an ultrasound machine, a serious shortcoming believed to have contributed to some of the unexpected results. These conclusions should come as no surprise to the creators of the HeartWorks simulator, and will hopefully be addresses in forthcoming releases, as the 3D model of the human heart that is the core of the simulator is without peer.

The switching-replications experimental design used worked very well and contributed greatly to the success of the project. By controlling for potentially strong
social effects that could have endangered internal validity, and also by maximizing the
data available for analysis, the switching replications design as used proved its worth.

Similarly, some of the constructivist-based features of the ID educational unit
used in the experimental phase resulted in positive results and feedback from participants. However, cautionary findings relating to the ID also included the need to carefully
evaluate the use of some features, as there was a tendency for participants to not value the
performance of certain features if they were not going to be graded, despite their possible educational benefit. The use of the simulator was embedded in the constructivist-guided
features of the educational unit as supporting structure for learning. Some of the features
of the educational unit were valued more than others by the participants, and for varying
reasons. One of the most valued constructivist features of the educational unit was the
scaffolding support process that was used to provide assistance to the participants in their
use of the simulator. Less well-received constructivist features of the educational unit
were the portfolio, Web quest, and learner reflective blog. These were almost universally
ignored in the experimental phase by each group, an outcome foreshadowed and forecast
by the pilot run-through earlier in the Spring. As in the pilot, the nearly universal
response from participants as to why these activities were not used was that they were
perceived as extra work and were not valued unless they were going to be graded. These
features, especially the portfolio and Web quest, could easily be assessed for a grade if
the educational unit were embedded in a course. In fact, these constructivist ID elements
were designed with assessment in mind, if fully implemented into an actual course.
However, as described earlier and to avoid any difficulties of undue influence to a potentially vulnerable population, it was communicated from the beginning that nothing the participants did or did not do would affect their grade in the actual host course for the experiment. The importance of performing these activities was however continually reinforced all participants during all interactions, the posting of reminders in the simulator room, and emails to group participants during their respective turns in the experimental unit. At least for these participants in this setting, the effect of not being evaluated for a grade was seemingly more powerful even than their volunteerism. This is an interesting outcome that would be worth exploring in a future, related study. It is also a cautionary note to others who may wish to implement what may be thought of as worthwhile but possibly time-consuming ID features into their own studies, as this effect seemed to overpower even the otherwise excellent participatory attitudes of the participants. The time-spent perception combined with the no grade effect, seemed to be a powerful set of influences against the performance of some activities.

Features of the educational unit that received modest support were the weekly image-based quiz and the discussion board. In contrast to the preceding discussion of the no-grade effect, the relative acceptance of the quizzes illustrated the fine line between the need to assess learner performance and the ability to provide effective learning in order to achieve desired educational effects. The discussion board was used by a few students to provide feedback and suggestions throughout the experiment. Although participation was low overall, thoughtful comments and suggestion were made by some participants and in
some cases acted upon. However, the anticipated asynchronous dialogue of sharing tips or helpful hints between learners as they moved through the educational unit never occurred. This could be an effect of the limited duration of only four weeks of each phase of the experiment. Perhaps an experimental phase of longer duration, or one that did not have the same level of strong scaffolding support for the learners, would have generated more discussion traffic of the anticipated nature, as the discussions then would have filled a need for helpful information.

Another takeaway for the ID structure and features of the educational unit was that the scaffolding support required by the learners was not nearly as complex in content-level, nor as time-intense, and did not need to remain at a high level of support for as long as originally planned. This could have been a result of the advance preparation in the form of the basic training on the simulator that was held for all students prior to the experiment. The additional support available in the form of quick-reference sheets, online tutorials, and the availability context-specific help within the simulator itself may have also reduced the learners’ need for intensive and extensive scaffolding support.

One of the biggest factors contributing to the success of the educational unit was that the participants were excited, engaged, and eager to learn on the simulator. This type of human factor may be hard to achieve and influence in all cases. However, much effort was expended to foster excitement and interest, including the initial briefing over pizza lunch, the award at the end of gift cards, continual positive reinforcement during
scaffolding sessions, and so on. All of the preceding discussion points are based upon comments made by participants during the experimental phase of the study and anecdotal observations gathered during informal Q&A sessions at the end of each four-week phase of the experiment.

**Future research.** Future research may include repetition of the basic study in other settings involving larger numbers of participants, repetition across similar institutions with geographically and demographically disparate student populations, and use of the educational unit ID template to implement simulation technology into other educational realms. Other possible extensions suggested include further research to determine the effects of a more-realistic UI design and/or additional life-like features in the simulation itself on learning outcomes. Additional qualitative research could include end-of-experimental-phase structured interviews to determine participant satisfaction and learning outlooks, and investigation of the learners’ thoughts and perceptions as they use actual ultrasound machines after practicing on the simulator, perhaps through think-aloud and active interview techniques.
Appendix A

Comprehensive Four Week Long Lesson Plan

Describes an educational unit integrating simulator technology using scaffolding, self-reflection, social learning, multiple learning pathways, and mastery learning techniques.

<table>
<thead>
<tr>
<th>Weekly Learning Focus</th>
<th>Student Activities</th>
<th>Instructor Activities and Supportive Resources</th>
</tr>
</thead>
</table>
| Week 1: Review of heart anatomy     | **Day 1:** (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment; possible additional time as needed).  
• Establish Simulator profile.  
• Use Simulator in Anatomy mode to review major external and internal structures of heart anatomy.  
• Create initial learner blog entry at end of session, self-reflecting on day’s learning (required).  
• Create initial entry in Bb peer Discussion group.  
• Select favorite anatomical feature of the heart as the focus of the Web Quest Assignment in Blackboard.  
• View supportive Tegrity and PowerPoint Course Content in the Bb course as needed during and after each Day’s Simulator session. | **Instructor:** facilitates all Day 1 activities in Week 1.  
**Resources for Week 1:**  
• Simulator in Anatomy mode.  
• Simulator room workstation.  
• Lync setup from Simulator room workstation to Instructor’s workstation.  
• Blackboard (Bb) course resources:  
  1. Learner-focused Blog and Discussion board (facilitated).  
  2. Tegrity sessions available for review of simulator setup and operation.  
  3. Course Content of PowerPoint presentations with embedded instructional videos on simulator use in Anatomy mode.  
  4. Weekly activities for the Web Quest Bb Assignment administered and managed through the course. |
| **Day 2**: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment) | **Instructor**: initially facilitates simulator activities in Day 2 for the 1st half-hour, and then checks in on the half-hour, or as needed in response to Lync request for assistance.  
Instructor discusses learner progress at end of daily session with learner.  
Instructor reviews individual Blog postings, moderates Discussion board postings, and ensures learners have selected a Web Quest Assignment.  
Instructor assists if needed in learner’s initial efforts to capture simulator images for Web Quest and portfolio use, discusses progress. |
| --- | --- |
| • Continue using Simulator in Anatomy mode to review major external and internal structures of heart anatomy.  
• Capture views of favorite anatomical feature of the heart for Web Quest Assignment and for learner portfolio.  
• Post self-reflective blog entry at end of Day (optional).  
• Participate in Bb peer Discussion group.  
• Perform visually-based unit formative assessment quiz at end of Day 2. |  
**Instructor**: ensures learner gets started in Day 3, checks in on the hour, and otherwise provides assistance in response to Lync requests only.  
Instructor discusses and reviews learner’s progress in Web Quest Assignment at end of Day.  
**AS APPLICABLE**: Instructor may discuss learner self-assessed mastery of sub-unit and administer and review summative test results with learner to jointly decide learner’s course of action. |
| **Day 3**: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment) |  
• Continue using Simulator in Anatomy mode to review major external and internal structures of heart anatomy.  
• Capture views of favorite anatomical feature of the heart for Web Quest Assignment and for learner portfolio.  
• Post self-reflective blog entry (optional).  
• Participate in Discussion board.  
• Perform visually-based unit formative assessment quiz at end of Day 3.  
• **OPTIONAL**: If learner self-assesses mastery of sub-unit, the summative assessment test may |
be taken and if passed with at least 80% correct, the learner may advance to the next sub-unit of instruction.

<table>
<thead>
<tr>
<th><strong>Day 4:</strong> (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment).</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Continue using Simulator in Anatomy mode to review major external and internal structures of heart anatomy.</td>
</tr>
<tr>
<td>• Capture views of favorite anatomical feature of the heart for Web Quest and portfolio.</td>
</tr>
<tr>
<td>• Self-reflective blog entry (optional).</td>
</tr>
<tr>
<td>• Participate in Discussion board.</td>
</tr>
<tr>
<td>• Perform visually-based unit formative assessment quiz at end of Day 4.</td>
</tr>
<tr>
<td>• End of Day 4 simulator session: the learner reviews progress and performance in daily formative quizzes, for self-reflective attainment of mastery. Discusses with Instructor, and has the option to continue the sub-unit for up to two more days, or move on to Day 5 and the next sub-unit of instruction.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Instructor:</strong> ensures learner gets started in Day4, checks in on the hour, and otherwise provides assistance in response to Lync requests only.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor reviews/moderates and provides feedback on Blog, Discussion board and Web Quest learner activities.</td>
</tr>
<tr>
<td>Reviews learner performance on formative quizzes in the unit, discusses with learner to decide next steps in sub-unit for the course.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Day 5:</strong> (Activity duration= 1 hours on simulator, 2 hours on Blackboard)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Continue using Simulator in Anatomy mode to review major external and internal structures of heart anatomy.</td>
</tr>
<tr>
<td>• Capture views of favorite anatomical feature of the heart for Web Quest and portfolio.</td>
</tr>
<tr>
<td>• Participate in Discussion board.</td>
</tr>
<tr>
<td>• Perform visually-based unit summative assessment test at end of Day 5.</td>
</tr>
<tr>
<td>• Post final Blog entry for the Week, self-reflecting on the</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Instructor:</strong> For Day 5, ensures learner gets started, and then provides assistance in response to Lync requests only.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor reviews learner Blog and Discussion board entries, reviews and provides feedback on Part I of Web Quest Assignment.</td>
</tr>
<tr>
<td>Instructor administers visually-based unit summative assessment test at end of Day 5 session, reviews and discusses results with learner. If needed,</td>
</tr>
</tbody>
</table>
week’s learning activities. (This is the only other required self-reflection besides Day 1, although daily entries are encouraged).
- Submit Part I of the Web Quest Assignment in Blackboard.
- Complete download of favorite anatomical images for use in learner portfolio and Web Quest.

<table>
<thead>
<tr>
<th>Weekly Learning Focus</th>
<th>Student Activities</th>
<th>Instructor Activities and Supportive Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week 2: Simulator-assisted practice of 6 standard imaging planes in Dual View mode (both anatomical and ultrasound views in use).</strong></td>
<td><strong>Day 1: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment; possible additional time as needed).</strong></td>
<td>Instructor: facilitates all Day 1 activities in Week 2.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Resources for Week 2:</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Simulator in Dual View mode.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Simulator room workstation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Lync setup from Simulator room workstation to Instructor’s workstation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Blackboard (Bb) course resources: Same as Week 1 with additional Course Content of Tegrity sessions and PowerPoint presentations with embedded instructional videos on the use of the Simulator in Dual View mode.</td>
</tr>
<tr>
<td></td>
<td><strong>Day 2: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)</strong></td>
<td>Instructor: initially facilitates simulator activities in Day 2 for the 1st half-hour, and then checks in on the half-hour, or as needed in response to Lync request for assistance.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Continue using Simulator in Dual View mode to master the 6</td>
</tr>
</tbody>
</table>
standard imaging planes of the heart.

- Capture Dual View images for use in Web Quest Assignment and in learner portfolio.
- Post self-reflective blog entry at end of Day (optional).
- Participate in Bb peer Discussion group.
- Perform visually-based unit formative assessment quiz at end of Day 2.

Instructor discusses learner progress at end of daily session with learner.

**Day 3:** (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)

- Continue using Simulator in Dual View mode to master the 6 standard imaging planes of the heart.
- Capture Dual View images for use in Web Quest Assignment and in learner portfolio.
- Post self-reflective blog entry at end of Day (optional).
- Participate in Bb peer Discussion group.
- Perform visually-based unit formative assessment quiz at end of Day 3.
- OPTIONAL: If learner self-assesses mastery of sub-unit, the summative assessment test may be taken and if passed with at least 80% correct, the learner may to advance to the next sub-unit of instruction.

Instructor ensures learner gets started in Day 3, checks in on the hour, and otherwise provides assistance in response to Lync requests only.

Instructor discusses and reviews learner’s progress in Web Quest Assignment at end of Day.

Instructor discusses learner progress at end of daily session with learner.

AS APPLICABLE: Instructor may discuss learner self-assessed mastery of sub-unit and administer and review summative test results with learner to jointly decide learner’s course of action.

**Day 4:** (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment).

- Continue using Simulator in Dual View mode to master the 6 standard imaging planes of the heart.

Instructor ensures learner gets started in Day4, checks in on the hour, and otherwise provides assistance in response to Lync requests only.

Instructor reviews/moderates and provides feedback on Blog.
<table>
<thead>
<tr>
<th>Day 4:</th>
<th></th>
<th>Day 5:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Capture Dual View images for use in Web Quest Assignment and in learner portfolio.</td>
<td>- Discussion board and Web Quest learner activities.</td>
<td>- Instructor: For Day 5, ensures learner gets started, and then provides assistance in response to Lync requests only.</td>
</tr>
<tr>
<td>- Post self-reflective blog entry at end of Day (optional).</td>
<td>- Reviews learner performance on formative quizzes in the unit, discusses with learner to decide next steps in sub-unit for the course.</td>
<td>Instructor reviews learner Blog and Discussion board entries, reviews and provides feedback on Part II of the Web Quest Assignment.</td>
</tr>
<tr>
<td>- Participate in Bb peer Discussion group.</td>
<td>- Perform visually-based unit formative assessment quiz at end of Day 4.</td>
<td>Instructor administers visually-based unit summative assessment test at end of Day 5 session, reviews and discusses results with learner. If needed, up to two additional days are added to ensure mastery of sub-unit topics.</td>
</tr>
<tr>
<td>- Perform visually-based unit formative assessment quiz at end of Day 4.</td>
<td>- End of Day 4 simulator session: the learner reviews progress and performance in daily formative quizzes, for self-reflective attainment of mastery. Discusses with Instructor, and has the option to continue the sub-unit for up to two more days, or move on to Day 5 and the next sub-unit of instruction.</td>
<td></td>
</tr>
<tr>
<td>- Day 5: (Activity duration= 1 hours on simulator, 2 hours on Blackboard)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Continue using Simulator in Dual View mode to master the 6 standard imaging planes of the heart.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Capture Dual View images for use in Web Quest Assignment and in learner portfolio.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Participate in Bb peer Discussion group.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Perform visually-based unit summative assessment test at end of Day 5.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Post final Blog entry for the Week, self --reflecting on the week’s activities. (This is the only other required self-reflection besides Day 1, although daily entries are encouraged).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Submit Part II of the Web Quest Assignment in Blackboard.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Complete download of favorite anatomical images for use in learner portfolio and Web Quest.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Learning Focus</td>
<td>Student Activities</td>
<td>Instructor Activities and Supportive Resources</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------</td>
<td>-----------------------------------------------</td>
</tr>
</tbody>
</table>
| **Week 3: Simulator-assisted practice of 6 standard imaging planes in Ultrasound mode only** | **Day 1**: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment; possible additional time as needed).  
- Use Simulator to master the 6 standard imaging planes of the heart in Ultrasound view mode.  
- Continue learner blog entries with self-reflection on Day 1 of Week 3’s learning activities (required).  
- Continue participation in Bb peer Discussion group.  
- Capture Ultrasound images for use in Web Quest Assignment and in learner portfolio.  
- View supportive Tegrity and PowerPoint Course Content in the Bb course as needed during and after each Day’s Simulator session.  | **Instructor**: facilitates all Day 1 activities in Week 3.  
**Resources for Week 3:**  
- Simulator in Ultrasound mode.  
- Simulator room workstation.  
- Lync setup from Simulator room workstation to Instructor’s workstation.  
- Blackboard (Bb) course resources: Same as Weeks 1 & 2 with additional Course Content of Tegrity sessions and PowerPoint presentations with embedded instructional videos on the use of the Simulator in Ultrasound mode.  |
|                      | **Day 2**: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)  
- Continue using Simulator in Ultrasound mode to master the 6 standard imaging planes of the heart.  
- Capture Ultrasound images for use in Web Quest Assignment and in learner portfolio.  
- Post self-reflective blog entry at end of Day (optional).  
- Participate in Bb peer Discussion group.  
- Perform visually-based unit formative assessment quiz at end of Day 2.  | **Instructor**: initially facilitates simulator activities in Day 2 for the 1st half-hour, and then checks in on the half-hour, or as needed in response to Lync request for assistance.  
Instructor reviews individual Blog postings, moderates Discussion board postings, and ensures learners have made progress on Part III of the Web Quest Assignment.  
Instructor discusses learner progress at end of daily session with learner.  |
### Day 3: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)

- Continue using Simulator in Ultrasound mode to master the 6 standard imaging planes of the heart.
- Capture Ultrasound images for use in Web Quest Assignment and in learner portfolio.
- Post self-reflective blog entry at end of Day (optional).
- Participate in Bb peer Discussion group.
- Perform visually-based unit formative assessment quiz at end of Day 3.
- **OPTIONAL:** If learner self-assesses mastery of sub-unit, the summative assessment test may be taken and if passed with at least 80% correct, the learner may to advance to the next sub-unit of instruction.

### Instructor:

- Ensures learner gets started in Day 3, checks in on the hour, and otherwise provides assistance in response to Lync requests only.
- Discusses learner progress at end of daily session with learner.
- Discusses and reviews learner’s progress in Web Quest Assignment at end of Day.
- **AS APPLICABLE:** May discuss learner self-assessed mastery of sub-unit and administer and review summative test results with learner to jointly decide learner’s course of action.

### Day 4: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)

- Continue using Simulator in Ultrasound mode to the 6 standard imaging planes of the heart.
- Capture Ultrasound images for use in Web Quest Assignment and in learner portfolio.
- Post self-reflective blog entry at end of Day (optional).
- Participate in Bb peer Discussion group.
- Perform visually-based unit formative assessment quiz at end of Day 4.
- **End of Day 4 simulator session:** The learner reviews progress and performance in daily formative quizzes, for self-reflective

### Instructor:

- Ensures learner gets started in Day 4, checks in on the hour, and otherwise provides assistance in response to Lync requests only.
- Reviews learner performance on formative quizzes in the unit, discusses with learner to decide next steps in sub-unit for the course.
- Reviews/moderates and provides feedback on Blog, Discussion board and Web Quest learner activities.
attainment of mastery. Discusses with Instructor, and has the option to continue the sub-unit for up to two more days, or move on to Day 5 and the next sub-unit of instruction.

**Day 5:** (Activity duration= 1 hours on simulator, 2 hours on Blackboard)

- Continue using Simulator in Ultrasound mode to master the 6 standard imaging planes.
- Capture Ultrasound images for use in Web Quest Assignment and in learner portfolio.
- Participate in Bb peer Discussion group.
- Perform visually-based unit summative assessment test at end of Day 5.
- Post final Blog entry for the Week, self-reflecting on the week’s activities. (This is the only other required self-reflection besides Day 1, although daily entries are encouraged).
- Submit Part III of the Web Quest Assignment in Blackboard.
- Complete download of favorite anatomical images for use in learner portfolio and Web Quest.

**Instructor:** For Day 5, ensures learner gets started, and then provides assistance in response to Lync requests only.

Instructor reviews learner Blog and Discussion board entries, reviews and provides feedback on Part III of the Web Quest Assignment.

Instructor administers visually-based unit summative assessment test at end of Day 5 session, reviews and discusses results with learner.

If needed, up to two additional days are added to ensure mastery of sub-unit topics.

<table>
<thead>
<tr>
<th>Weekly Learning Focus</th>
<th>Student Activities</th>
<th>Instructor Activities and Supportive Resources</th>
</tr>
</thead>
</table>
| **Week 4:** Use of the simulator’s Comparative feature to review and then assess the ability of the student to capture the 6 standard planes, versus the ideal standard approach | **Day 1:** (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment; possible additional time as needed).  
- Use Simulator to master the 6 standard imaging planes of the heart in Comparative view mode.  
- Continue learner blog entries with self-reflection on Day 1 of Week 4’s learning activities (required). | **Instructor:** facilitates all Day 1 activities in Week 4.  
**Resources for Week 4:**  
- Simulator in Comparative mode.  
- Simulator room workstation.  
- Lync setup from Simulator room workstation to Instructor’s workstation.  
- Blackboard (Bb) course resources: |
<table>
<thead>
<tr>
<th>Day 1:</th>
<th>Same as Weeks 1, 2 &amp; 3 with additional Course Content of Tegrity sessions and PowerPoint presentations with embedded instructional videos on the use of the Simulator in Comparative mode.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continue participation in Bb peer Discussion group.</td>
<td>Instructor: initially facilitates simulator activities in Day 2 for the 1st half-hour, and then checks in on the half-hour, or as needed in response to Lync request for assistance.</td>
</tr>
<tr>
<td>Capture Comparative view mode images for use in Web Quest Assignment and in learner portfolio.</td>
<td>Instructor reviews individual Blog postings, moderates Discussion board postings, and ensures learners have made progress on Part IV of the Web Quest Assignment.</td>
</tr>
<tr>
<td>View supportive Tegrity and PowerPoint Course Content in the Bb course as needed during and after each Day’s Simulator session.</td>
<td>Instructor discusses learner progress at end of daily session with learner.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 2: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)</th>
<th>Instructor: ensures learner gets started in Day 3, checks in on the hour, and otherwise provides assistance in response to Lync requests only.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continue using Simulator in Comparative mode to master the 6 standard imaging planes of the heart.</td>
<td>Instructor discusses learner progress at end of daily session with learner.</td>
</tr>
<tr>
<td>Capture Comparative images for use in Web Quest Assignment and in learner portfolio.</td>
<td>Instructor discusses and reviews learner’s progress in Web Quest Assignment at end of Day.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 3: (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Continue using Simulator in Comparative mode to master the 6 standard imaging planes of the heart.</td>
<td></td>
</tr>
<tr>
<td>Capture Comparative images for use in Web Quest Assignment and in learner portfolio.</td>
<td></td>
</tr>
<tr>
<td>Post self-reflective blog entry at end of Day (optional).</td>
<td></td>
</tr>
</tbody>
</table>
- Participate in Bb peer Discussion group.
- Perform visually-based unit formative assessment quiz at end of Day 3.
- **OPTIONAL:** If learner self-assesses mastery of sub-unit, the summative assessment test may be taken and if passed with at least 80% correct, the learner may to advance to the next sub-unit of instruction.

**AS APPLICABLE:** Instructor may discuss learner self-assessed mastery of sub-unit and administer and review summative test results with learner to jointly decide learner’s course of action.

<table>
<thead>
<tr>
<th><strong>Day 4:</strong> (Activity duration= 1 hours on simulator, 1 hour on Blackboard, 1 hour on Web Quest Assignment)</th>
<th><strong>Instructor:</strong> ensures learner gets started in Day 4, checks in on the hour, and otherwise provides assistance in response to Lync requests only. Instructer reviews/moderates and provides feedback on Blog, Discussion board and Web Quest learner activities. Reviews learner performance on formative quizzes in the unit, discusses with learner to decide next steps in sub-unit for the course.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Continue using Simulator in Comparative mode to master the 6 standard imaging planes of the heart.</td>
<td></td>
</tr>
<tr>
<td>- Capture Comparative view mode images for use in Web Quest Assignment and in learner portfolio.</td>
<td></td>
</tr>
<tr>
<td>- Post self-reflective blog entry at end of Day (optional).</td>
<td></td>
</tr>
<tr>
<td>- Participate in Bb peer Discussion group.</td>
<td></td>
</tr>
<tr>
<td>- Perform visually-based unit formative assessment quiz at end of Day 4.</td>
<td></td>
</tr>
<tr>
<td>- End of Day 4 simulator session: the learner reviews progress and performance in daily formative quizzes, for self-reflective attainment of mastery. Discusses with Instructor, and has the option to continue the sub-unit for up to two more days, or move on to Day 5 and the next sub-unit of instruction.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Day 5:</strong> (Activity duration= 1 hours on simulator, 2 hours on Blackboard)</th>
<th><strong>Instructor:</strong> For Day 5, ensures learner gets started, and then provides assistance in response to Lync requests only.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Continue using Simulator in Comparative mode to master the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>6 standard imaging planes of the heart.</td>
<td></td>
</tr>
<tr>
<td>Capture Comparative view mode images for use in Web Quest Assignment and in learner portfolio.</td>
<td></td>
</tr>
<tr>
<td>Participate in Bb peer Discussion group.</td>
<td></td>
</tr>
<tr>
<td>Perform visually-based unit summative assessment test at end of Day 5.</td>
<td></td>
</tr>
<tr>
<td>Post final Blog entry for the entire Instructional Unit (required).</td>
<td></td>
</tr>
<tr>
<td>Submit Part IV of the Web Quest Assignment in Blackboard.</td>
<td></td>
</tr>
<tr>
<td>Complete download of favorite anatomical images for use in learner portfolio and Web Quest.</td>
<td></td>
</tr>
<tr>
<td>Complete and submit Web Quest Assignment via Blackboard.</td>
<td></td>
</tr>
<tr>
<td>Organize final selected images for submission of learner imaging portfolio.</td>
<td></td>
</tr>
</tbody>
</table>

Instructor reviews learner Blog and Discussion board entries, reviews and provides feedback on Part IV of the Web Quest Assignment.

Instructor administers visually-based unit summative assessment test at end of Day 5 session, reviews and discusses results with learner.

If needed, up to two additional days are added to ensure mastery of sub-unit topics.
Appendix B

Focused Three Week Long Lesson Plan

Used for the pilot implementation of the experimental educational unit, plan uses scaffolding, self-reflection, social learning, multiple learning pathways, and mastery learning techniques.

<table>
<thead>
<tr>
<th>Weekly Learning Focus</th>
<th>Student Activities</th>
<th>Instructor Activities and Supportive Resources</th>
</tr>
</thead>
</table>
| **Week 1: Simulator-assisted practice of 6 targeted standard imaging approaches in Dual View mode (both anatomical and ultrasound views in use).** | **Day 1:** (Activity duration 2 hours = 1 hour on simulator, 1 hour on Blackboard and Web Quest Assignment; possible additional time as needed).  
- Use Simulator in Dual View mode to view and practice the targeted standard imaging planes of the heart.  
- Begin learner blog entries with self-reflection after Day 1 of Week 1’s learning activities (required).  
- Begin participation in Bb peer Discussion group.  
- Capture Dual View images for use in Web Quest Assignment and in learner portfolio.  
- As needed, view supportive Tegrity and PowerPoint Course Content in the Bb course as needed during and after each Simulator session. | **Instructor:** facilitates all Day 1 activities in Week 1. Instructor discusses learner progress at end of all daily sessions with learner.  
**Resources for Week 1:**  
- Simulator in Dual View mode, 1 hour per student.  
- Simulator room workstation.  
- Lync setup from Simulator room workstation to Instructor’s workstation.  
- Blackboard (Bb) course resources:  
  1. Learner-focused Blog and Discussion board (facilitated).  
  2. Tegrity sessions available for review of simulator profile setup and operation.  
  3. Course Content of PowerPoint presentations with embedded instructional videos on simulator use.  
  4. Weekly activities for the Web Quest |
### Weekly Learning Focus

| Week 2: Simulator-assisted practice of 6 targeted standard imaging planes in Ultrasound mode only. |

### Student Activities

| Day 1: (Activity duration 2 hours = 1 hour on simulator, 1 hour on Blackboard and Web Quest Assignment; possible additional time as needed). |

### Instructor Activities and Supportive Resources

| Assignment administered and managed through the Bb course.  
5. Image-based formative quizzes are hosted and administered in the Bb course.  

Instructor: For Day 2 Week 1, instructor facilitates simulator activities for first 1/2 hour, then checks in at end of 1-hour session, or as needed is response to Lync request for assistance.  

Instructor reviews learner Blog and Discussion board entries, reviews and provides feedback on Part II of the Web Quest Assignment.  

Instructor administers visually-based unit formative assessment test at end of Day 2 session, reviews and discusses results with learner.  

If needed and agreed upon with the learner, up to two additional 1-hour sessions may be added to ensure mastery of sub-unit topics.  

- Continue using Simulator in Dual View mode to master the targeted standard imaging planes of the heart.  
- Capture Dual View images for use in Web Quest Assignment and in learner portfolio.  
- Participate in Bb peer Discussion group.  
- Perform visually-based unit formative assessment quiz at end of Day 2.  
- Post Blog entry for Week 1, self-reflecting on the week’s activities. (This is the only other required self-reflection in Week 1 besides Day 1, although entries are encouraged at any time).  
- Submit Part I of the Web Quest Assignment in Blackboard.  
- Complete download of captured images for use in learner portfolio and Web Quest.  

Instructor: facilitates all Day 1 activities in Week 2.  

Instructor discusses learner progress at end of all daily sessions with learner.
<table>
<thead>
<tr>
<th>NOTE: Each set of daily activities may occur on any calendar day during Week 2 based on scheduling and availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Use Simulator to master the targeted standard imaging planes of the heart in Ultrasound view mode.</td>
</tr>
<tr>
<td>• Continue learner blog entries with self-reflection on Day 1 of Week 3’s learning activities (required).</td>
</tr>
<tr>
<td>• Continue participation in Bb peer Discussion group.</td>
</tr>
<tr>
<td>• Capture Ultrasound images for use in Web Quest Assignment and in learner portfolio.</td>
</tr>
<tr>
<td>• View supportive Tegrity and PowerPoint Course Content in the Bb course as needed during and after each Day’s Simulator session.</td>
</tr>
<tr>
<td>OPTIONAL: If learner self-assesses mastery of sub-unit, the formative assessment test for the unit may be taken and if passed with at least 90% correct, the learner may advance to the next sub-unit of instruction.</td>
</tr>
<tr>
<td>Resources for Week 2:</td>
</tr>
<tr>
<td>• Simulator in Ultrasound mode.</td>
</tr>
<tr>
<td>• Simulator room workstation.</td>
</tr>
<tr>
<td>• Lync setup from Simulator room workstation to Instructor’s workstation.</td>
</tr>
<tr>
<td>• Blackboard (Bb) course resources:</td>
</tr>
<tr>
<td>Same as Week 1 with additional Course Content of Tegrity sessions and PowerPoint presentations with embedded instructional videos on the use of the Simulator in Ultrasound mode.</td>
</tr>
<tr>
<td>AS APPLICABLE: Instructor may discuss learner self-assessed mastery of sub-unit and administer and review summative test results with learner to jointly decide learner’s course of action.</td>
</tr>
<tr>
<td>Day 2: (Activity duration 2 hours = 1 hours on simulator, 1 hour on Blackboard and Web Quest Assignment; possible additional time as needed).</td>
</tr>
<tr>
<td>• Continue using Simulator in Ultrasound mode to master the targeted standard imaging planes of the heart.</td>
</tr>
<tr>
<td>• Capture Ultrasound images for use in Web Quest Assignment and in learner portfolio.</td>
</tr>
<tr>
<td>• Post self-reflective blog entry at end of Week 2 (optional).</td>
</tr>
<tr>
<td>• Participate in Bb peer Discussion group.</td>
</tr>
<tr>
<td>Instructor: initially facilitates simulator activities in Day 2 for the 1st half-hour, and then at end of session or as needed in response to Lync request for assistance.</td>
</tr>
<tr>
<td>Instructor reviews individual Blog postings, moderates Discussion board postings, and ensures learners have made progress on Part II of the Web Quest Assignment.</td>
</tr>
<tr>
<td>Instructor administers visually-based unit summative assessment test at end of Day 5</td>
</tr>
</tbody>
</table>
**Weekly Learning Focus**

**Week 3: Use of the simulator’s Comparative feature to review and then assess the ability of the student to capture the 6 targeted standard imaging planes, versus the ideal standard approach portrayed by the simulator**

**NOTE:** Each set of daily activities may occur on any calendar day during Week 3 based on scheduling and availability.

---

**Student Activities**

**Day 1:** (Activity duration 2 hours = 1 hour on simulator, 1 hour on Blackboard and Web Quest Assignment; possible additional time as needed).

- Perform visually-based unit formative assessment quiz by end of Week 2.
- Submit Part II of the Web Quest Assignment by end of Week 2.
- Learner reviews progress and performance for self-reflective attainment of mastery. Discusses with Instructor, and has the option to continue the sub-unit for up to two more days.
- Use Simulator to master the targeted standard imaging planes of the heart in Comparative view mode.
- Continue learner blog entries with self-reflection on Day 1 of Week 4’s learning activities (required).
- Continue participation in Bb peer Discussion group.
- Capture Comparative view mode images for use in Web Quest Assignment and in learner portfolio.
- View supportive Tegrity and PowerPoint Course Content in the Bb course as needed during and after each Day’s Simulator session.

**Instructor Activities and Supportive Resources**

**Instructor:** facilitates all Day 1 activities in Week 3.

**Resources for Week 3:**

- Simulator in Comparative mode.
- Simulator room workstation.
- Lync setup from Simulator room workstation to Instructor’s workstation.
- Blackboard (Bb) course resources: Same as Weeks 1, 2 & 3 with additional Course Content of Tegrity sessions and PowerPoint presentations with embedded instructional videos on the use of the Simulator in Comparative mode.

Instructor discusses learner progress at end of daily session with learner.
OPTIONAL: If learner self-assesses mastery of sub-unit, the formative assessment test for the unit may be taken and if passed with at least 90% correct, the learner may advance to the next sub-unit of instruction.

AS APPLICABLE: Instructor may discuss learner self-assessed mastery of sub-unit and administer and review summative test results with learner to jointly decide learner’s course of action.

**Day 2:** (Activity duration 2 hours = 1 hour on simulator, 1 hour on Blackboard and Web Quest Assignment; possible additional time as needed).

- Continue using Simulator in Comparative mode to master the targeted standard imaging planes of the heart.
- Capture Comparative images for use in Web Quest Assignment and in learner portfolio.
- Post self-reflective blog entry at end of Day (optional).
- Participate in Bb peer Discussion group.
- Perform visually-based unit formative assessment quiz at end of Day 2.
- End of Day 2 simulator session: the learner reviews progress and performance in daily formative quizzes for self-reflective attainment of mastery. Discusses with Instructor, and has the option to continue the sub-unit for up to two more days.

Instructor: initially facilitates simulator activities in Day 2 for the 1st half-hour, and then checks in at the end of the session, or as needed in response to Lync request for assistance.

Instructor reviews individual Blog postings, moderates Discussion board postings, and ensures learners have made progress on Part III of the Web Quest Assignment.

Instructor discusses learner progress at end of session with learner.

If needed, and mutually agreed upon, up to two additional 1-hour sessions are added to ensure mastery of sub-unit topics.
## Appendix C

### Standardized Assessment Tool for Transthoracic Echocardiography (TTE)

<table>
<thead>
<tr>
<th>Study ID:</th>
<th>Evaluation Date:</th>
<th>Circle: Simulated or Patient TTE</th>
<th>Evaluator’s Initials:</th>
</tr>
</thead>
</table>

### 1. PLAX View

- **View is Correct, Check**
  - Yes
  - No

**Check Identifiable Structures**
- Inferior Vena Cava
- Right Atrium

**SA Score:** \_/2

### 2. RVIT View

- **View is Correct, Check**
  - Yes
  - No

**Check Identifiable Structures**
- Left Ventricle
- Right Ventricle
- Left Atrium
- Right Atrium
- Tricuspid Valve
- Mitral Valve
- InterventricularSeptum

**SA Score:** \_/7

### 3. RVOT View

- **View is Correct, Check**
  - Yes
  - No

**Check Identifiable Structures**
- Right Ventricle
- Pulmonary Artery
- Pulmonic Valve

**SA Score:** \_/3

### 4. PSAX View, AV Level

- **View is Correct, Check**
  - Yes
  - No

**Check Identifiable Structures**
- Left Ventricle
- Mitral Valve
- Right Ventricle

**SA Score:** \_/7

### 5. PSAX View, MV Level

- **View is Correct, Check**
  - Yes
  - No

**Check Identifiable Structures**
- Left Ventricle
- Mitral Valve
- Right Ventricle
- Interventricular Septum

**SA Score:** \_/4

### 6. PSAX View, Papillary Level

- **View is Correct, Check**
  - Yes
  - No

**Check Identifiable Structures**
- Left Ventricle
- Papillary Muscles

**SA Score:** \_/2

### BTTEET Full Exam, Duration (if applicable)

- **Minutes:** 
- **Seconds:**

**Views correct:** \_/6

**SA Total:** \_/35

**KEY:** TTE = Transthoracic Echocardiography, IVC= Inferior vena Cava, SA = Structural Anatomy, Subc. = Subcostal, SAX = Short Axis View, AV = Aortic Valve, MV = Mitral Valve, Prox. = Proximal CCA = Common Carotid Artery, BTTEET = Basic Transthoracic Echocardiography Evaluation Tool.
### Appendix D

**Standardized Assessment Exam for Clinical Evaluation of Echocardiography**

**Mid-term practicum: Echo I CVS 3010**

Name: ________________________ Date: _______

1. You will scan two patients.
2. On the first patient, you will acquire one clip and one frozen image for each of the 6 views: PLAX, RVIT, RVOT, PSAX AV, PSAX MV, and PSAX Papillary. In your frozen image for each view, label all of the structures listed in the word-bank below that are seen in that view. You will have 30 minutes.

   Patient 1 name: ___________________________________________
   Exam Room: ____________ Start time: ________ Finish time: ______

3. On the second patient, you will acquire one clip and one frozen image for each of the 6 views. In your frozen image for each view, label all of the structures listed in the word-bank below that are seen in that view. You will have 30 minutes.

   Patient 2 name: ___________________________________________
   Exam Room: ____________ Start time: ________ Finish time: ______

Grading: You start with 100 points; detailed rubric available for review upon request.

   - _______ Accuracy of view (60 points); I am looking to see if you show the anatomy required, if the scan plane is correct, etc.
   - _______ Appropriate labelling of the following ten anatomical structures in every frozen view in which they appear. Acceptable abbreviations are listed: (30 points)

   1. Right Ventricle (RV)
   2. Left Ventricle (LV)
   3. Tricuspid Valve (TV)
   4. Interventricular Septum (IVS)
   5. Pulmonary Artery (PA)
   6. Pulmonic Valve (PV)
   7. Aortic Valve (AV)
   8. Left Atrium (LA)
   9. Right Atrium (RA)
   10. Papillary Muscles (Pap.)
   11. Aortic Root (Ao)
   12. Mitral Valve (MV)

   _______ Appropriate gains, depth, & other image quality issues (10 points)

Comments:
Appendix E

MEMORANDUM

To: Samuel Yoders, EdS
    College of Engineering and Computing

From: Ling Wang, Ph.D.,
    Center Representative, Institutional Review Board

Date: November 17, 2015

Re: IRB #: 2015-125; Title, “Integration of Simulation into Healthcare Education through Applied Constructivism: A Switching Replications Randomized Controlled Trial”

I have reviewed the above-referenced research protocol at the center level. Based on the information provided, I have determined that this study is exempt from further IRB review under 45 CFR 46.101(b) (Exempt Category 1). You may proceed with your study as described to the IRB. As principal investigator, you must adhere to the following requirements:

1) CONSENT: If recruitment procedures include consent forms, they must be obtained in such a manner that they are clearly understood by the subjects and the process affords subjects the opportunity to ask questions, obtain detailed answers from those directly involved in the research, and have sufficient time to consider their participation after they have been provided this information. The subjects must be given a copy of the signed consent document, and a copy must be placed in a secure file separate from de-identified participant information. Record of informed consent must be retained for a minimum of three years from the conclusion of the study.

2) ADVERSE EVENTS/UNANTICIPATED PROBLEMS: The principal investigator is required to notify the IRB chair and me (954-262-5369 and Ling Wang, Ph.D., respectively) of any adverse reactions or unanticipated events that may develop as a result of this study. Reactions or events may include, but are not limited to, injury, depression as a result of participation in the study, life-threatening situation, death, or loss of confidentiality/anonymity of subject. Approval may be withdrawn if the problem is serious.

3) AMENDMENTS: Any changes in the study (e.g., procedures, number or types of subjects, consent forms, investigators, etc.) must be approved by the IRB prior to implementation. Please be advised that changes in a study may require further review depending on the nature of the change. Please contact me with any questions regarding amendments or changes to your study.

Cc: Gertrude Abramson, Ed.D.
Appendix F

Random Sequence Generator Used to Randomly Assign Participants to Experimental Groups

Volunteers were assigned to the experimental group (Group I) or the initial control group (Group II) using the number sequence generated paired to their consent form number.

RANDOM.ORG - Sequence Generator
Random Sequence Generator

This form allows you to generate randomized sequences of integers. The randomness comes from atmospheric noise, which for many purposes is better than the pseudo-random number algorithms typically used in computer programs.

Part 1: Sequence Boundaries
Smallest value 1 (limit -1,000,000,000)
Largest value 18 (limit+1,000,000,000)
Format in 2 column(s)
The length of the sequence (the largest minus the smallest value plus 1) can be no greater than 10,000.

Part 2: Go!
Be patient! It may take a little while to generate your sequence...
Get Sequence Reset Form
Switch to Advanced Mode

Note: A randomized sequence does not contain duplicates (the numbers are like raffle tickets drawn from a hat). There is also the Integer Generator which generates the numbers independently of each other (like rolls of a die) and where each number can occur more than once.

Random Sequence Generator
Here is your sequence:

12 7
4 8
11 2
18 3
14 16
1 15
13 6
Again!  Go Back

Note: The numbers are generated left to right, i.e., across columns.
Follow @RandomOrg (3,691 followers)

© 1998-2016 RANDOM.ORG
Terms and Conditions
About Us
Appendix G

SPSS Normality Testing and Subsequent Parametric Analysis of Pilot Data

SPSS version 24.0.0.0 output for normality testing and the subsequent parametric statistical analysis via independent sample t-tests comparing performance criteria in the pilot run-through between the 10 student volunteers versus the 12 abstaining students in the pilot run-through. Parametric analysis performed using time-to-complete data and standardized clinical assessment scores post-treatment for the volunteer participants, normal learning path for abstaining students.

Normality testing of baseline scores for both groups; if normality of distribution is satisfied, parametric testing can be used.

<table>
<thead>
<tr>
<th>Clinical Combined Score (SA &amp; Views), Both Groups</th>
<th>Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>23.2273</td>
<td>1.28078</td>
</tr>
<tr>
<td>95% Confidence Interval for Mean: Lower Bound</td>
<td>20.5637</td>
<td></td>
</tr>
<tr>
<td>95% Confidence Interval for Mean: Upper Bound</td>
<td>25.8908</td>
<td></td>
</tr>
<tr>
<td>5% Trimmed Mean</td>
<td>23.3636</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>23.0000</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>36.089</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>6.00739</td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>11.00</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>33.00</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>22.00</td>
<td></td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>7.75</td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.541</td>
<td>0.491</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.412</td>
<td>0.953</td>
</tr>
</tbody>
</table>

The numerical measurements of skewness and kurtosis provide for an objective numbers-based investigation of normality (Park, 2008). The Descriptives shown above for this case list a negative skewness value of -0.541 and negative kurtosis at -0.412. These values mean, respectively, that the observed distribution occurs slightly more-frequently to the left of, or less-than the mean value (negative skewness) and has lower values than...
would be expected in a strictly normal distribution thus resulting in a more-flattened observed distribution (platykurtosis) (Terrell, 2012, pp. 105-107). For skewness, a normal distribution would have a value of zero, and skewness values less than or equal to -2 or +2 indicate a non-significant variation from a normal distribution (Terrell, 2012, p. 106). In our case negative skewness is indicated by the skewness value of -0.541, but as it is still greater than negative two the observed distribution is not significantly different from a normal distribution. Similarly, and since SPSS actually reports kurtosis-minus-three (Park, 2008, p.8), a kurtosis value less than zero indicates negative kurtosis or platykurtosis, seen as a lower peak with thicker or taller tails if the distribution is plotted (Park, 2008, p.37). Kurtosis values greater than positive 2 or less than negative 2 are indicative of a significantly non-normal distribution for the observed data (Terrell, 2012, p. 106). Although platykurtosis of the observed distribution is indicated by the negative kurtosis value of -0.412, it is still greater than negative two and thus not significantly different from a normal distribution.

### Tests of Normality

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Combined Score (SA &amp; Views), Both Groups</td>
<td>.192 22 .034</td>
<td>.937 22 .171</td>
</tr>
</tbody>
</table>

a. Lilliefors Significance Correction

The statistical tests of normality provide for objective, numerical examination of normality of distribution. The Shapiro-Wilk test of normality should be used when \( n = 4 \) to 2000. For Shapiro-Wilk the test hypothesis is that the observed distribution is non-normal (does not fit a normal distribution). The null test hypothesis is that the observed distribution is normal (fits a normal distribution) (Park, 2008, p. 9). In this case the value of \( p = .171 \) is greater than our alpha value of 0.05, meaning the test is non-significant, so we fail to reject the null test hypothesis, with the conclusion that the distribution fits a normal distribution, or, is relatively normal. Shown in the Statistic column for the Shapiro-Wilk is the W statistic which is always less than or equal to positive one. A W statistic close to one indicates normal distribution (Park, 2008, p. 8). In this case, a W statistic of 0.937 is shown, indicating relative normality. Therefore, both parameters of the Shapiro-Wilk test indicate that the observed distribution is relatively normal.

Graphical methods of investigating normality provide descriptive, visually intuitive means of investigating normality of distribution (Park, 2008, p. 3). The histogram provides a descriptive means of subjectively evaluating the relative normality of a data distribution.
In this case, although the missing values in the lower range results in a gap in the pattern, the histogram shows a relatively normal distribution that has thicker tails than a normal distribution (corresponding to the numerical finding of platykurtosis) and more heavily represented to the left side of the mean (corresponding to the numerical finding of negative positive skewness). In this case there were many values at 22 and 23, just below the mean of 23.23; with a slightly different mean based upon the visual appearance of the histogram of this distribution it would likely have evaluated as having positive skewness instead. However, the overall shape still conforms to a relatively normal distribution with no secondary or tertiary peaks (bi-modal or tri-modal distribution).
The Normal Q-Q Plot provides another visual means of subjectively evaluating the relative normality of an observed data distribution (Park, 2008, p. 6). The Q-Q Plot in this case shows minimal variation from the expected normal values if the data are normally distributed (y-axis) versus the observed values (x-axis) with regularity of plotted values seen along the central line of normality and no clustering of values plotted. Normality testing statement: Since normality of distribution was satisfied with consistent results from both objective and subjective methods, parametric testing may be used (Park, 2008, p. 36).

Independent sample t –test for significant positive difference in combined standardized assessment scores between the volunteer students (group 1, \(n = 10\)) versus the abstaining students (group 2, \(n = 12\)). Test hypothesis (one-tailed, directional): Student volunteers (group 1) scored significantly higher assessment scores compared to the abstaining students (expected result). Null test hypothesis: There was no significant difference in assessment scores between the two groups of students.

NOTE: Comparison using a combined clinical score consisting of Structural Anatomy (SA) score and correct views score added together.

### T-Test

<table>
<thead>
<tr>
<th></th>
<th>Group Statistics</th>
<th></th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group, either</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volunteers (V) or</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstainers (A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Combined Score</td>
<td>Volunteers</td>
<td>10</td>
<td>27.0000</td>
<td>3.62093</td>
</tr>
<tr>
<td>Score (SA &amp; Views), Both</td>
<td>Abstainers</td>
<td>12</td>
<td>20.0833</td>
<td>5.86915</td>
</tr>
<tr>
<td>Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Independent Samples Test

<table>
<thead>
<tr>
<th>Clinical Combined Score</th>
<th>Equal variances</th>
<th>Equal variances not</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>assumed</td>
<td>assumed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levene's Test for Equality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>4.742</td>
<td>.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-test for Equality of</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>3.241</td>
<td>3.382</td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>20</td>
<td>18.600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.004</td>
<td>.003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of $p$ for Levene’s is .042; since this is less than .05, equal variances are not assumed. That is, we reject the null hypothesis comparing the variances between the two groups, meaning that there is a significant difference in variance in the scores between the two groups, and we must use the Equal Variances Not Assumed column for the analysis of the independent sample $t$-test.

Analysis using the Equal Variances Not Assumed column: The computed $p$ value (Sig. 2-tailed) is .003; with a one-tailed hypothesis this value must be divided by two. The one-tailed $p$ value of .0015 is much less than our alpha value of .05, indicating significance, so we must decide to reject the null test hypothesis.

We consider the direction of the means with our mean values of 27.0000 for the volunteers group and 20.0833 for the abstainers. Therefore, we have a finding that there are higher average combined scores for the volunteer group compared to the average combined scores for the abstainers. This supports the test hypotheses and the decision to reject the null test hypothesis.

The critical value of $t$ for $df = 20$ at $\alpha = .05$ is 1.725. With the computed value of $t$ at 3.241 being greater than the critical value of $t$ we must support the test hypothesis and decide to reject the null hypothesis that there was no significant difference in scores between groups.

The decision to reject the null hypothesis is supported by the large effect size found. Computation of Cohen’s delta ($d$) for effect size for the independent sample $t$–test:

\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}, \text{ where}
\]

\[
S_{\text{pooled}} = \sqrt{\frac{\text{standard deviation}_1^2 + \text{standard deviation}_2^2}{2}}
\]

\[
S_{\text{pooled}} = \sqrt{\frac{(3.62093^2 + 5.86915^2)/2}{2}}
\]

\[
S_{\text{pooled}} = \sqrt{\frac{(13.11113 + 34.44692)/2}{2}}
\]

\[
S_{\text{pooled}} = \sqrt{\frac{(47.55805)/2}{2}}
\]

\[
S_{\text{pooled}} = \sqrt{23.77902}
\]

\[
S_{\text{pooled}} = 4.87637
\]

\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}
\]
\[ d = (27.0000 - 20.0833)/4.8763 \]
\[ d = 6.9167/4.8763 \]
\[ d = 1.42 \]

This is a very large effect size, e.g. the level of the independent variable (membership in either the volunteer group, or in the abstaining group) has a very large effect on the dependent variable (clinical combined assessment score) and supports the decision to reject the null test hypothesis. Effect size descriptions as provided by Cohen (1988), and Sawilowsky (2009):

- Very small, \( d = 0.01 \)
- Small, \( d = 0.20 \)
- Medium, \( d = 0.50 \)
- Large, \( d = 0.80 \)
- Very large, \( d = 1.20 \)
- Huge, \( d = 2.0 \) or higher

Findings statement for independent sample \( t \)–test for the existence of a significant positive difference in combined clinical assessment scores with the volunteer group higher than the abstaining group:

On average, the volunteer students scored significantly higher in combined clinical assessment scores than the abstaining students; \( t (20) = 3.241, p = .0015 \) at \( \alpha = .05 \)

Independent sample \( t \)–test for significant positive difference in standardized assessment scores expressed as a percentage, between the volunteer students (group V, \( n = 10 \)) versus the abstaining students (group A, \( n = 12 \)). Test hypothesis (one-tailed, directional):
Student volunteers (group V) scored significantly higher assessment scores expressed as a percentage, when compared to the abstaining students (expected result).
Null test hypothesis: There was no significant difference in assessment score percentages between the two groups of students.
NOTE: Comparison using combined clinical score consisting of Structural Anatomy (SA) score and number of correct views score added together, expressed as a percentage.

**T-Test**

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Volunteers (V)</th>
<th>Abstainers (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Volunteers</td>
<td>10</td>
<td>.8200</td>
</tr>
</tbody>
</table>

| 154 |
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of $p$ for Levene’s is .030; since this is less than .05, equal variances are not assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is a significant difference in variance in the scores between the two groups when expressed as a percentage, and we must use the Equal Variances Not Assumed column for the analysis of the independent sample $t$-test.

Analysis using the Equal Variances Not Assumed column: The computed $p$ value (Sig. 2-tailed) is .003; with a one-tailed hypothesis this value must be divided by two. The one-tailed $p$ value of .0015 is much less than our alpha value of .05, indicating significance, so we must decide to reject the null test hypothesis.

We consider the direction of the means with our mean values of .8200 (or 82%) for the volunteers group and .6083 (or 60.83%) for the abstainers. Therefore, we have a finding that there are higher average combined scores expressed as a percentage for the volunteer

<table>
<thead>
<tr>
<th>Clinical Combined Score as a Percentage, Both Groups</th>
<th>Abstainers</th>
<th>12</th>
<th>.6083</th>
<th>.18080</th>
<th>.05219</th>
</tr>
</thead>
</table>

**Independent Samples Test**

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>F</th>
<th>5.482</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig.</td>
<td></td>
<td>.030</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-test for Equality of Means</td>
<td>t</td>
<td>3.239</td>
<td>3.388</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>20</td>
<td>18.379</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.004</td>
<td>.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Difference</td>
<td></td>
<td>.21167</td>
<td>.21167</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Error Difference</td>
<td></td>
<td>.06534</td>
<td>.06248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% Confidence Interval of the Difference</td>
<td>Lower</td>
<td>.07537</td>
<td>.08059</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>.34797</td>
<td>.34274</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
group compared to the average combined scores percentage for the abstainers. This supports the test hypotheses and the decision to reject the null test hypothesis.

The critical value of $t$ for $df = 20$ at $\alpha = .05$ is 1.725. With the computed value of $t$ at 3.239 being greater than the critical value of $t$ we must support the test hypothesis and decide to reject the null hypothesis that there was no significant difference in score percentages between groups.

The decision to reject the null hypothesis is supported by the large effect size found. Computation of Cohen’s delta ($d$) for effect size for the independent sample $t$–test:

\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}, \text{ where} \\
S_{\text{pooled}} = \sqrt{\frac{S_{1}^2 + S_{2}^2}{2}} \\
S_{\text{pooled}} = \sqrt{\frac{(.108632 + .180802)/2}{2}} \\
S_{\text{pooled}} = \sqrt{\frac{(.01180 + .03268)/2}{2}} \\
S_{\text{pooled}} = \sqrt{\frac{(.04448)/2}{2}} \\
S_{\text{pooled}} = \sqrt{.02224} \\
S_{\text{pooled}} = .14913 \\
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}} \\
d = \frac{(.8200 - .6083)}{.14913} \\
d = .2117 / .14913 \\
d = 1.42
\]
This is a very large effect size, e.g. the level of the independent variable (membership in either the volunteer group, or in the abstaining group) has a very large effect on the dependent variable (clinical combined assessment score percentage) and supports the decision to reject the null test hypothesis.

Findings statement for independent sample $t$–test for the existence of a significant positive difference in combined clinical assessment score percentage with the volunteer group higher than the abstaining group:
On average, the volunteer students had significantly higher combined clinical assessment percent scores than the abstaining students; $t (20) = 3.239, p = .0015$ at $\alpha = .05$

Independent sample $t$–test for significant positive difference in standardized assessment number of correct clinical views between the volunteer students (group 1, $n = 10$) versus the abstaining students (group 2, $n = 12$). Test hypothesis (one-tailed, directional):
Student volunteers (group 1) had significantly higher assessment scores for the number of correct views when compared to the abstaining students (expected result).
Null test hypothesis: There was no significant difference in assessment scores for the number of correct views between the two groups of students.

**T-Test**

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Group, either</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volunteers</td>
<td>10</td>
<td>4.5000</td>
<td>1.17851</td>
<td>.37268</td>
<td></td>
</tr>
<tr>
<td>Abstainers</td>
<td>12</td>
<td>3.3333</td>
<td>1.49747</td>
<td>.43228</td>
<td></td>
</tr>
</tbody>
</table>

**Independent Samples Test**

<table>
<thead>
<tr>
<th>Levene’s Test for Equality of Variances</th>
<th>Equal variances assumed</th>
<th>Equal variances not assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F</strong></td>
<td>.280</td>
<td></td>
</tr>
<tr>
<td><strong>Sig.</strong></td>
<td>.603</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-test for Equality of Means</th>
<th>Clinical Views Score, Both Groups</th>
<th>Equal variances assumed</th>
<th>Equal variances not assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T</strong></td>
<td></td>
<td>1.999</td>
<td>2.044</td>
</tr>
<tr>
<td><strong>Df</strong></td>
<td></td>
<td>20</td>
<td>19.955</td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td></td>
<td>.059</td>
<td>.054</td>
</tr>
<tr>
<td><strong>Mean Difference</strong></td>
<td></td>
<td>1.16667</td>
<td>1.16667</td>
</tr>
<tr>
<td><strong>Std. Error Difference</strong></td>
<td></td>
<td>.58369</td>
<td>.57075</td>
</tr>
<tr>
<td><strong>95% Confidence Interval of the Difference</strong></td>
<td>Lower</td>
<td>-.05089</td>
<td>-.02407</td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>2.38422</td>
<td>2.35741</td>
</tr>
</tbody>
</table>

Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of $p$ for Levene’s is .603; since this is greater than or equal to .05, equal variances are assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is no significant difference in variance in the scores between the two groups, and we can use the Equal Variances Assumed column for the analysis of the independent sample $t$-test.

Analysis using the Equal Variances Assumed column: The computed $p$ value (Sig. 2-tailed) is .059; with a one-tailed hypothesis this value must be divided by two. The one-tailed $p$ value of .0295 is less than our alpha value of .05, indicating significance, so we must decide to reject the null test hypothesis.
We consider the direction of the means with our mean values of 4.5000 for the volunteers group and 3.3333 for the abstainers. Therefore, we have a finding that there are higher average number of correct views scores for the volunteer group compared to the average correct views scores for the abstainers. This supports the test hypotheses and a decision to reject the null test hypothesis.

The critical value of $t$ for $df = 20$ at $\alpha = .05$ is 1.725. With the computed value of $t$ at 1.999 being greater than the critical value of $t$ we must accept the test hypothesis and decide to reject the null hypothesis that there was no significant difference in number of correct views scores between groups.

The decision to accept the test hypothesis and reject the null hypothesis is supported by the large effect size found. Computation of Cohen’s delta ($d$) for effect size for the independent sample $t$-test:

$$d = \frac{\text{Mean Difference}}{S_{\text{pooled}}},$$

where

$S_{\text{pooled}} = \sqrt{\frac{(1.17851^2 + 1.49747^2)/2}{2}}$

$S_{\text{pooled}} = \sqrt{\frac{((1.38888 + 2.24241)/2)}{2}}$

$S_{\text{pooled}} = \sqrt{\frac{(3.63129)/2}{2}}$

$S_{\text{pooled}} = \sqrt{((1.81564)/2)}$

$S_{\text{pooled}} = 1.34745$

$d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}$

$d = (4.5000 - 3.3333)/ 1.34745$

$d = 1.1667 / 1.41973$

$d = 0.822$

This is a large effect size, e.g. the level of the independent variable (membership in either the volunteer group, or in the abstaining group) has a large effect on the dependent variable (clinical correct views score) and supports the decision to accept the test hypothesis and reject the null hypothesis.

Findings statement for independent sample $t$-test for the existence of a significant positive difference in clinical assessment correct views scores with the volunteer group higher than the abstaining group:

On average, the volunteer students scored significantly higher in clinical assessment number of correct views than the abstaining students; $t (20) = 1.999, p = .0295$ at $\alpha = .05$
Independent sample $t$–test for a significant positive difference in standardized assessment structural anatomy (SA) scores between the volunteer students (group 1, $n = 10$) versus the abstaining students (group 2, $n = 12$).

Test hypothesis (one-tailed, directional): Student volunteers (group 1) scored significantly higher assessment scores compared to the abstaining students (group 2), (expected result).

Null test hypothesis: There was no significant difference in SA assessment scores between the two groups of students.

**T-Test**

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Group, either Volunteers (V) or Abstainers (A)</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical SA Score, Both Groups</td>
<td>Volunteers</td>
<td>10</td>
<td>22.5000</td>
<td>2.50555</td>
<td>.79232</td>
</tr>
<tr>
<td></td>
<td>Abstainers</td>
<td>12</td>
<td>16.7500</td>
<td>4.65393</td>
<td>1.34347</td>
</tr>
</tbody>
</table>

**Independent Samples Test**

<table>
<thead>
<tr>
<th>Levene’s Test for Equality of Variances</th>
<th>F</th>
<th>.023</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-test for Equality of Means</td>
<td>t</td>
<td>3.498</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.002</td>
</tr>
<tr>
<td>Mean Difference</td>
<td>5.75000</td>
<td>5.75000</td>
</tr>
<tr>
<td>Std. Error Difference</td>
<td>1.64374</td>
<td>1.55971</td>
</tr>
<tr>
<td>95% Confidence Interval of the Difference</td>
<td>Lower</td>
<td>2.32122</td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>9.17878</td>
</tr>
</tbody>
</table>

Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of $p$ for Levene’s is .023; since this is less than .05, equal variances are not assumed. That is, we must accept the null hypothesis comparing the variances between the two groups, meaning that there a significant difference in variance in the scores between the two
groups, and we must use the Equal Variances Not Assumed column for the analysis of the independent sample $t$-test.

Analysis using the Equal Variances Not Assumed column: The computed $p$ value (Sig. 2-tailed) is .002; with a one-tailed hypothesis this value must be divided by two. The one-tailed $p$ value of .001 is much less than our alpha value of .05, indicating significance, so we must decide to reject the null test hypothesis.

We consider the direction of the means with our mean values of 22.50 for the volunteers group and 16.75 for the abstainers. Therefore, we have a finding that there are higher average SA scores for the volunteer group compared to the average SA scores for the abstainers. This supports the test hypotheses and the decision to reject the null test hypothesis.

The critical value of $t$ for $df = 20$ at $\alpha = .05$ is 1.725. With the computed value of $t$ at 3.498 being greater than the critical value of $t$ we must support the test hypothesis and decide to reject the null hypothesis that there was no significant difference in scores between groups.

The decision to reject the null hypothesis is supported by the large effect size found.

Computation of Cohen’s delta ($d$) for effect size for the independent sample $t$-test:

\[
d = \frac{\text{Mean Difference}}{\text{S}_\text{pooled}}, \text{ where}
\]

$\text{S}_\text{pooled} = \sqrt{\text{standard deviation}_1^2 + \text{standard deviation}_2^2}/2$

$\text{S}_\text{pooled} = \sqrt{(2.50555^2 + 4.65393^2)/2}$

$\text{S}_\text{pooled} = \sqrt{(6.27778 + 21.65906)/2}$

$\text{S}_\text{pooled} = \sqrt{(27.93684)/2}$

$\text{S}_\text{pooled} = \sqrt{13.96842}$

$\text{S}_\text{pooled} = 3.73743$

\[
d = \frac{4.65393 - 2.50555}{3.73743}
\]

\[
d = 2.14838 / 3.73743
\]

\[
d = 0.57
\]

This is a large effect size, e.g. the level of the independent variable (membership in either the volunteer group, or in the abstaining group) has a large effect on the dependent variable (clinical assessment SA score) and supports the decision to reject the null test hypothesis.

Findings statement for independent sample $t$-test for the existence of a significant positive difference in clinical assessment SA scores with the volunteer group higher than the abstaining group:
On average, the volunteer students scored significantly higher in clinical assessment structural anatomy scores than the abstaining students; \( t (20) = 3.498, p = .001 \) at \( \alpha = .05 \)

Independent sample \( t \)-test for differences in total time to complete the six views of the clinical assessment between the volunteer students versus the students who abstained. Test hypothesis (two-tailed, non-directional): there was a significant difference in total time to complete the six clinical assessment views between the volunteer and abstaining students. Null test hypothesis: There was no significant difference in total time to complete the six clinical views between the two groups.

**T-Test**

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Group, either Volunteers (V) or Abstainers (A)</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Total Time to Obtain all 6 Views, Both Groups</td>
<td>Volunteers</td>
<td>10</td>
<td>671.3000</td>
<td>201.57825</td>
<td>63.74464</td>
</tr>
<tr>
<td></td>
<td>Abstainers</td>
<td>12</td>
<td>679.5000</td>
<td>307.19183</td>
<td>88.67864</td>
</tr>
</tbody>
</table>

Independent Samples Test

<table>
<thead>
<tr>
<th></th>
<th>Clinical Total Time to Obtain all 6 Views, Both Groups</th>
<th>( F )</th>
<th>( \text{Sig.} )</th>
<th>( \text{df} )</th>
<th>( \text{df} ) for Equality of Means</th>
<th>( \text{Sig. (2-tailed)} )</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levene's Test for Equality of Variances</td>
<td>Equal variances assumed</td>
<td>.686</td>
<td>.417</td>
<td>20</td>
<td>19.079</td>
<td>.943</td>
<td>941</td>
<td>-8.20000</td>
<td>113.43548</td>
</tr>
<tr>
<td>t-test for Equality of Means</td>
<td>Equal variances not assumed</td>
<td>-0.72</td>
<td>-0.75</td>
<td>20</td>
<td>19.079</td>
<td>.943</td>
<td>941</td>
<td>-8.20000</td>
<td>113.43548</td>
</tr>
</tbody>
</table>
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of \( p \) for Levene’s is .417 (> .05), so equal variances are assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is no significant difference in variance between the two groups, and we can use the Equal Variances Assumed column for the analysis of independent sample \( t \)-test.

Analysis using the Equal Variances Assumed column: The computed \( p \) value (Sig. 2-tailed) of .943 is much greater than our alpha value of .05 divided by two, equals .025 (for the non-directional test hypothesis). Therefore, we must decide to not reject the null hypothesis (fail to reject).

The critical value of \( t \) for \( df = 20 \) at \( \alpha = .025 \) is 2.086 (+, -). With the computed value of \( t \) at -.072 within the range of the critical value of \( t \) at + or -2.086 we must fail to reject the null hypothesis that there was no significant difference in baseline scores between groups.

The decision to fail to reject the null hypothesis is supported by the very small effect size found. Computation of Cohen’s delta (\( d \)) for effect size for the independent sample \( t \)-test:

\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}, \text{ where}
\]
\[
S_{\text{pooled}} = \text{square root of standard deviation}_1 \text{ squared, plus standard deviation}_2 \text{ squared, divided by 2}, \text{ or}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{((201.57825^2 + 307.19183^2)/2)}{2}}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{((40633.79087 + 94366.82042)/2)}{2}}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{((135000.6113)/2)}{2}}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{(67500.30565)}{2}}
\]
\[
S_{\text{pooled}} = 259.80820
\]
\[
d = \frac{\text{Mean Difference} / S_{\text{pooled}}}{d = (679.5000 – 671.3000) / 259.80820}
\]
\[
d = 8.2000 / 259.80820
\]
\[
d = 0.031
\]

This is a very small effect size; that is, the level of the independent variable (membership in either the volunteer group, or in the abstaining group) has a very small effect on the dependent variable (average time to complete the six clinical assessment views) and supports the decision to fail to reject the null test hypothesis.

Findings statement for independent sample \( t \)-test for the existence of a significant difference in total time to complete the six views in the clinical assessment between the volunteer group and the abstaining group:

On average, there was no significant difference in the total time to complete the six clinical views between groups; \( t (20) = -.072, p = .943 \) at \( \alpha = .05 \).
Independent sample $t$-test for differences in average time to obtain the six views (average per-view time) of the clinical assessment between the volunteer students versus the students who abstained.

Test hypothesis (two-tailed, non-directional): there was a significant difference in average time to obtain each of the six clinical assessment views between the volunteer and abstaining students. Null test hypothesis: There was no significant difference between the two groups in average time to obtain each of the six clinical views.

**T-Test**

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Group, either Volunteers (V) or Abstainers (A)</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Average Time to Obtain Views, Both Groups</td>
<td>Volunteers</td>
<td>10</td>
<td>111.8820</td>
<td>33.59700</td>
<td>10.62430</td>
</tr>
<tr>
<td></td>
<td>Abstainers</td>
<td>12</td>
<td>113.2508</td>
<td>51.19833</td>
<td>14.77968</td>
</tr>
</tbody>
</table>

**Independent Samples Test**

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.686</td>
<td>.417</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-test for Equality of Means</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.072</td>
<td>20</td>
<td>.943</td>
<td>-.136883</td>
<td>18.90592</td>
<td>-40.80589 - 39.45551</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>.941</th>
<th>19.079</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Difference</td>
<td>-1.36883</td>
<td>-1.36883</td>
</tr>
<tr>
<td>Std. Error Difference</td>
<td>18.90592</td>
<td>18.20206</td>
</tr>
<tr>
<td>95% Confidence Interval of the Difference</td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>-40.80589</td>
<td>38.06823</td>
<td>-39.45551</td>
</tr>
</tbody>
</table>
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of $p$ for Levene’s is .417 (> .05), so equal variances are assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is no significant difference in variance between the two groups, and we can use the Equal Variances Assumed column for the analysis of independent sample $t$-test.

Analysis using the Equal Variances Assumed column: The computed $p$ value (Sig. 2-tailed) of .943 is much greater than our alpha value of .05 divided by two, equals .025 (for the non-directional test hypothesis). Therefore, we must decide to not reject the null hypothesis (fail to reject).

The critical value of $t$ for $df = 20$ at $\alpha = .025$ is 2.086 (+, -). With the computed value of $t$ at -.072 within the range of the critical value of $t$ at + or -2.086 we must fail to reject the null hypothesis that there was no significant difference in baseline scores between groups.

The decision to fail to reject the null hypothesis is supported by the very small effect size found. Computation of Cohen’s delta ($d$) for effect size for the independent sample $t$–test:

\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}} \text{, where}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{(\text{standard deviation}_1)^2 + (\text{standard deviation}_2)^2}{2}},
\]
divided by 2, or
\[
S_{\text{pooled}} = \sqrt{\frac{((33.59700)^2 + (51.19833)^2)}{2}}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{((1128.75840 + 2621.26899)/2)}{2}}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{((3750.02739)/2)}{2}}
\]
\[
S_{\text{pooled}} = \sqrt{\frac{(1875.01369)/2)}{2}}
\]
\[
S_{\text{pooled}} = 43.30142
\]
\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}
\]
\[
d = \frac{113.2508 - 111.8820}{43.30142}
\]
\[
d = 0.031
\]

This is a very small effect size; that is, the level of the independent variable (membership in either the volunteer group, or in the abstaining group) has a very small effect on the dependent variable (average time to obtain each of the six clinical assessment views) and supports the decision to fail to reject the null test hypothesis.

Findings statement for independent sample $t$ –test for the existence of a significant difference in average time to obtain each of the six views in the clinical assessment between student in the volunteer group and the abstaining group:

On average, there was no significant difference between groups in the average time to obtain each of the six clinical views; $t (20) = -.072$, $p = .943$ at $\alpha = .05$
Appendix H

SPSS Normality Testing and Subsequent Parametric Analysis of Experimental Data

SPSS version 24.0.0.0 output for normality testing and subsequent parametric analysis using independent sample t-tests for comparison of assessment scores between experimental and control groups. Parametric analysis performed using standardized assessment scores at baseline, post-treatment Round One, and post-treatment Round Two with switching replications experimental design group assignments.

Normality testing of baseline scores for both groups; if normality of distribution is satisfied, parametric testing can be used.

Explore

Descriptives

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>28.5556</td>
</tr>
<tr>
<td>Std. Error</td>
<td>2.96518</td>
</tr>
<tr>
<td>95% Confidence</td>
<td>22.2996</td>
</tr>
<tr>
<td>Interval for Mean</td>
<td>34.8115</td>
</tr>
<tr>
<td>Lower Bound</td>
<td>22.2996</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>34.8115</td>
</tr>
<tr>
<td>5% Trimmed Mean</td>
<td>28.2284</td>
</tr>
<tr>
<td>Median</td>
<td>26.0000</td>
</tr>
<tr>
<td>Variance</td>
<td>158.261</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>12.58020</td>
</tr>
<tr>
<td>Minimum</td>
<td>6.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>57.00</td>
</tr>
<tr>
<td>Range</td>
<td>51.00</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>20.75</td>
</tr>
<tr>
<td>Skewness</td>
<td>.405 .536</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>.020 1.038</td>
</tr>
</tbody>
</table>

Tests of Normality

<table>
<thead>
<tr>
<th>Kolmogorov-Smirnova</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>df</td>
</tr>
</tbody>
</table>
The numerical measurements of skewness and kurtosis provide for an objective numbers-based investigation of normality (Park, 2008). The Descriptives shown above for this case list a positive skewness value of 0.405 and positive kurtosis at 0.020. These values mean, respectively, that the observed distribution occurs more-frequently to the right of, or great-than the mean value (positive skewness) and has higher values thus resulting in a more-peaked observed distribution (leptokurtosis) than would be expected in a strictly normal distribution (Terrell, 2012, p. 107). For skewness, a normal distribution would have a value of zero, and skewness values less than or equal to -2 or +2 indicate a non-significant variation from a normal distribution (Terrell, 2012, p. 106). Similarly, and since SPSS actually reports kurtosis-minus-three (Park, 2008, p.8), a kurtosis value greater than zero indicates positive kurtosis or leptokurtosis, seen as a high peak with flat tails if the distribution is plotted, and kurtosis value less than zero indicates negative kurtosis or platykurtosis, seen as a flattened peak and taller, thicker tails if the distribution is plotted (Park, 2008, p.37). Kurtosis values greater than positive 2 or less than negative 2 are indicative of significantly non-normal distribution for the observed data (Terrell, 2012, p. 106).

The statistical tests of normality provide for objective, numerical examination of normality of distribution. The Shapiro-Wilk test of normality should be used when \( n = 4 \) to 2000 (Park, 2008, p. 9). For Shapiro-Wilk the test hypothesis is that the observed distribution does not fit normal distribution (is a non-normal distribution). The null test hypothesis is that the observed distribution fits a normal distribution (normality of distribution). In this case the value of \( p = .506 \) is greater than our alpha value of 0.05, meaning the test is non-significant, so we fail to reject the null test hypothesis, with the conclusion that the distribution is relatively normal. Shown in the Statistic column for the Shapiro-Wilk is the W statistic which is always less than or equal to positive one. A W statistic close to one indicates normal distribution (Park, 2008, p. 8). In this case, a W statistic of 0.955 is shown, indicating relative normality. Therefore, both parameters of the Shapiro-Wilk test indicate that the observed distribution is relatively normal.
Graphical methods of investigating normality provide descriptive, visually intuitive means of investigating normality of distribution (Park, 2008, p. 3). The histogram provides a descriptive means of subjectively evaluating the relative normality of a data distribution. In this case the histogram shows a relatively normal distribution that is peaked in shape (corresponding to the numerical finding of leptokurtosis) and more heavily represented to the right side of the mean (corresponding to the numerical finding of positive skewness). However, the overall shape still conforms to a relatively normal distribution with no secondary or tertiary peaks (bi-modal or tri-modal distribution).
The Normal Q-Q Plot provides another visual means of subjectively evaluating the relative normality of an observed data distribution (Park, 2008, p. 6). The Plot in this case shows minimal variation from the expected normal values if the data are normally distributed (y-axis) versus the observed values (x-axis) with regularity of plotted values seen along the central line of normality and no clustering of values plotted. Normality testing statement: Since normality of distribution was satisfied with consistent results from both objective and subjective methods, parametric testing may be used (Park, 2008, p. 36).

Independent sample $t$–test for significant difference in baseline assessment scores between the participants assigned to Group I versus the participants assigned to Group II.

Test hypothesis (two-tailed, non-directional): There was a significant difference in baseline assessment scores between the participants assigned to Group I and those assigned to Group II.

Null test hypothesis: There was no significant difference in baseline assessment scores between Groups (expected result).

**T-Test**

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Baseline Scores</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Group</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Scores</td>
<td>Group I</td>
<td>9</td>
<td>27.8889</td>
<td>10.09263</td>
</tr>
<tr>
<td></td>
<td>Group II</td>
<td>9</td>
<td>29.2222</td>
<td>15.27889</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Samples Test</th>
<th>Baseline Scores</th>
<th>Equal variances assumed</th>
<th>Equal variances not assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levene's Test for Equality of Variances</td>
<td>F</td>
<td>1.109</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.308</td>
<td></td>
</tr>
<tr>
<td>t-test for Equality of Means</td>
<td>t</td>
<td>-.218</td>
<td>-.218</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>16</td>
<td>13.865</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.830</td>
<td>.830</td>
</tr>
<tr>
<td>Mean Difference</td>
<td>-1.33333</td>
<td>-1.33333</td>
<td></td>
</tr>
<tr>
<td>Std. Error Difference</td>
<td>6.10378</td>
<td>6.10378</td>
<td></td>
</tr>
</tbody>
</table>
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of \( p \) for Levene’s is .308 (> .05), so equal variances are assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is no significant difference in variance between the two groups, and we can use the Equal Variances Assumed column for the analysis of independent sample \( t \)-test.

Analysis using the Equal Variances Assumed column: The computed \( p \) value (Sig. 2-tailed) of .830 is much greater than our alpha value of .05 divided by two, equals .025 (for the non-directional test hypothesis). Therefore, we must decide to not reject the null hypothesis (fail to reject).

The critical value of \( t \) for \( df = 16 \) at \( \alpha = .025 \) is 2.120 (+, -). With the computed value of \( t \) at -.218 within the range of the critical value of \( t \) at + or -2.120, we must fail to reject the null hypothesis that there was no significant difference in baseline scores between groups.

The decision to fail to reject the null hypothesis is supported by the small effect size found. Computation of Cohen’s delta (\( d \)) for effect size for the independent sample \( t \)–test:

\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}, \text{ where}
\]

\[
S_{\text{pooled}} = \sqrt{\frac{s_1^2 + s_2^2}{2}},
\]

\[
S_{\text{pooled}} = \sqrt{\frac{(10.09263^2 + 15.27889^2)/2}{2}}
\]

\[
S_{\text{pooled}} = \sqrt{\frac{101.861584 + 233.4444796}{2}}
\]

\[
S_{\text{pooled}} = \sqrt{\frac{335.3060630}{2}}
\]

\[
S_{\text{pooled}} = 12.94808989
\]

\[
d = \frac{-1.333}{12.94808989}
\]

\[
d = -0.102 \text{ (absolute value, since two-tailed) = .102}
\]

This is a small effect size, e.g. there is little effect on the dependent variable (the mean baseline scores of the two groups) by the independent variable (group assignment) and supports the decision to fail to reject the null hypothesis.

Findings statement for independent sample \( t \)–test for the existence of a significant difference in baseline scores between Group I and Group II:
On average, there was no significant difference in baseline assessment scores for participants in Group I versus participants in Group II; $t (16) = -0.218$, $p = .830$ at $\alpha = .05$.

Independent sample $t$-test for differences in Round One assessment scores between the participants in Group I (the initial experimental group) versus the participants assigned to Group II (the initial control group).

Test hypothesis (one-tailed, or directional): Students in Group I scored significantly higher in Round One assessment scores than participants in Group II (expected result). Null test hypothesis: There was no significant difference in Round One assessment scores between the two Groups.

**T-Test**

<table>
<thead>
<tr>
<th>Group Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned Random Group</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Round One Scores</td>
</tr>
<tr>
<td>Group I</td>
</tr>
<tr>
<td>Group II</td>
</tr>
</tbody>
</table>

**Independent Samples Test**

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>Equal variances assumed</th>
<th>Equal variances not assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>.003</td>
<td>.954</td>
</tr>
<tr>
<td>Sig.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-test for Equality of Means</th>
<th>Round One Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>3.469</td>
</tr>
<tr>
<td>df</td>
<td>15</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.003</td>
</tr>
</tbody>
</table>

| Mean Difference | 33.69444 | 33.69444 |
| Std. Error Difference | 9.71317 | 9.68748 |
| 95% Confidence Interval of the Difference | Lower | 12.99132 | 13.03351 |
|                                  | Upper | 54.39757 | 54.35538 |
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of $p$ for Levene’s is .954 (> .05), so equal variances are assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is no significant difference in variance in the scores between the two groups, and we can use the Equal Variances Assumed column for the analysis of independent sample t-test.

Analysis using the Equal Variances Assumed column: The computed $p$ value (Sig. 2-tailed) is .003; with a one-tailed hypothesis this value must be divided by two. The one-tailed $p$ value of .0015 is much less than our alpha value of .05, indicating significance, so we must decide to reject the null test hypothesis.

We consider the direction of the means with our mean values of 84.25 for Group I and 50.56 for Group II. Therefore, we have a finding that there are higher scores for Group I compared to scores for Group II in Round One of assessment, which supports the test hypotheses and the decision to reject the null test hypothesis.

The critical value of $t$ for $df = 15$ at $\alpha = .05$ is 1.753. With the computed value of $t$ at 3.469 greater than the critical value of $t$ we must support the decision to reject the null hypothesis that there was no significant difference in Round One scores between groups.

The decision to reject the null hypothesis is supported by the large effect size found.

Computation of Cohen’s delta ($d$) for effect size for the independent sample $t$–test:

\[
d = \frac{\text{Mean Difference}}{S_{\text{pooled}}}, \text{ where}
\]

\[
S_{\text{pooled}} = \sqrt{\text{standard deviation}_1^2 + \text{standard deviation}_2^2}, \text{ divided by 2, or}
\]

\[
S_{\text{pooled}} = \sqrt{((19.53568^2 + 20.37837^2)/2)}
\]

\[
S_{\text{pooled}} = \sqrt{((381.64279 + 415.27796)/2)}
\]

\[
S_{\text{pooled}} = \sqrt{((796.92075)/2)}
\]

\[
S_{\text{pooled}} = \sqrt{398.46037}
\]

\[
S_{\text{pooled}} = 19.96147
\]

\[
d = \frac{33.69444}{19.96147}
\]

\[
d = 1.68
\]

This is a very large effect size (> 1.20), e.g. the level of the independent variable (Group I or Group II) having a very large effect on the dependent variable (Round One assessment score) and supports the decision to reject the null test hypothesis.

Findings statement for independent sample $t$–test for the existence of a significant positive difference in Round One scores with Group I higher than Group II: On average, participants in Group I scored significantly higher in Round One assessment scores than participants in Group II; $t(15) = 3.469$, $p = .0015$ at $\alpha = .05$
Independent sample *t*-test for differences in Round Two assessment scores between the participants in Group I (the initial experimental group, the secondary control group in the switching replications design) versus the participants assigned to Group II (the initial control group, and secondary experimental group in the switching replications design).

Test hypothesis (two-tailed, non-directional): there was a significant difference in Round Two assessment scores between the participants in Group I and those in Group II.

Null test hypothesis: There was no significant difference in Round Two assessment scores between Groups (expected result).

### T-Test

#### Group Statistics

<table>
<thead>
<tr>
<th>Assigned Random Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Two Scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I</td>
<td>8</td>
<td>71.5000</td>
<td>14.36265</td>
<td>5.07796</td>
</tr>
<tr>
<td>Group II</td>
<td>9</td>
<td>83.7778</td>
<td>14.63538</td>
<td>4.87846</td>
</tr>
</tbody>
</table>

#### Independent Samples Test

<table>
<thead>
<tr>
<th></th>
<th>Round Two Scores</th>
<th>Equal variances assumed</th>
<th>Equal variances not assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levene’s Test for Equality of Variances</td>
<td>F</td>
<td>.198</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig.</td>
<td>.663</td>
<td></td>
</tr>
<tr>
<td><em>t</em>-test for Equality of Means</td>
<td>t</td>
<td>-1.742</td>
<td>-1.744</td>
</tr>
<tr>
<td></td>
<td>df</td>
<td>15</td>
<td>14.830</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.102</td>
<td>.102</td>
</tr>
<tr>
<td>Mean Difference</td>
<td>-12.27778</td>
<td>-12.27778</td>
<td></td>
</tr>
<tr>
<td>Std. Error Difference</td>
<td>7.04998</td>
<td>7.04167</td>
<td></td>
</tr>
<tr>
<td>95% Confidence Interval of the Difference</td>
<td>Lower</td>
<td>-27.30446</td>
<td>-27.30171</td>
</tr>
<tr>
<td></td>
<td>Upper</td>
<td>2.74891</td>
<td>2.74615</td>
</tr>
</tbody>
</table>
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of $p$ for Levene’s is .663 (> .05), so equal variances are assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is no significant difference in variance between the two groups, and we can use the Equal Variances Assumed column for the analysis of independent sample $t$-test.

Analysis using the Equal Variances Assumed column: The computed $p$ value (Sig. 2-tailed) of .102 is much greater than our alpha value of .05 divided by two, equals .025 (for the non-directional test hypothesis). Therefore, we must decide to not reject the null hypothesis (fail to reject).

The critical value of $t$ for $df = 15$ at $\alpha = .025$ is 2.131 (+, -). With the computed value of $t$ at -1.742 is within the range of the critical value of $t$ at + or - 2.131 we must fail to reject the null hypothesis that there was no significant difference in baseline scores between groups.

Interestingly, the decision to fail to reject the null hypothesis is not supported by the effect size found. Computation of Cohen’s delta ($d$) for effect size for the independent sample $t$–test:

$$d = \frac{\text{Mean Difference}}{\text{S}_{\text{pooled}}}, \text{ where}$$

$$\text{S}_{\text{pooled}} = \sqrt{\frac{\text{standard deviation}_1^2 + \text{standard deviation}_2^2}{2}},$$

$$\text{S}_{\text{pooled}} = \sqrt{\frac{(14.36265^2 + 14.63538^2)/2}{2}}$$
$$\text{S}_{\text{pooled}} = \sqrt{\frac{(206.28571 + 214.19434)/2}{2}}$$
$$\text{S}_{\text{pooled}} = \sqrt{\frac{(420.48005)/2}{2}}$$
$$\text{S}_{\text{pooled}} = \sqrt{210.24002}$$
$$\text{S}_{\text{pooled}} = 14.49946$$

$$d = \frac{-12.27778}{14.49946}$$
$$d = -0.84677 \text{ (absolute value) = .846}$$

This is a large effect size, e.g. there is a large effect on the dependent variable (the mean Round Two scores of the two groups) by the independent variable (group assignment) which is consistent with the observed difference between the group average means. However, the difference is not statistically significant, as described above.

Findings statement for independent sample $t$–test for the existence of a significant difference in Round Two scores between Group I and Group II:

On average, there was no significant difference in Round Two assessment scores for participants in Group I versus participants in Group II; $t (15) = -1.742, p = .102$ at $\alpha = .05$
Independent sample $t$-test for differences in Round Two for the time to complete all views between the participants in Group I (the initial experimental group, and the secondary control group in the switching replications design) versus the participants assigned to Group II (the initial control group, and secondary experimental group in the switching replications design).

Test hypothesis (two-tailed, non-directional): there was a significant difference in Round Two time to complete all views between the participants in Group I and those in Group II.

Null test hypothesis: There was no significant difference in Round Two time to complete all views between Groups (expected result).

### T-Test

<table>
<thead>
<tr>
<th>Assigned Random Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>8</td>
<td>37.2500</td>
<td>18.19537</td>
<td>6.43303</td>
</tr>
<tr>
<td>Group II</td>
<td>9</td>
<td>43.0000</td>
<td>23.58495</td>
<td>7.86165</td>
</tr>
</tbody>
</table>

**Independent Samples Test**

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.043</td>
<td>.173</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-test for Equality of Means</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.557</td>
<td>15</td>
<td>.586</td>
<td>-5.75000</td>
<td>10.32112</td>
<td>-27.74894</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean Difference: -5.75000
Std. Error Difference: 10.32112
95% Confidence Interval of the Difference: Lower -27.74894, Upper 16.24894
Levene’s Test for Equality of Variances: with our alpha level at .05, the computed value of \( p \) for Levene’s is .173 (> .05), so equal variances are assumed. That is, we fail to reject the null hypothesis comparing the variances between the two groups, meaning that there is no significant difference in variance between the two groups, and we can use the Equal Variances Assumed column for the analysis of independent sample \( t \)-test.

Analysis using the Equal Variances Assumed column: The computed \( p \) value (Sig. 2-tailed) of .586 is much greater than our alpha value of .05 divided by two, equals .025 (for the non-directional test hypothesis). Therefore, we must decide to not reject the null hypothesis.

The critical value of \( t \) for \( df = 15 \) at \( \alpha = .025 \) is 2.131 (+, -). With the computed value of \( t \) at -.557 within the range of the critical value of \( t \) at + or -2.120 we must fail to reject the null hypothesis that there was no significant difference in time to complete between groups.

The decision to fail to reject the null hypothesis is supported by the very small effect size found. Computation of Cohen’s delta (\( d \)) for effect size for the independent sample \( t \) – test:

\[
\begin{align*}
\text{d} &= \text{Mean Difference} / S_{\text{pooled}} \quad \text{where} \\
S_{\text{pooled}} &= \text{square root of standard deviation}_1 \text{ squared, plus standard deviation}_2 \text{ squared, divided by 2, or} \\
S_{\text{pooled}} &= \text{square root of} ( ((18.19537^2 + 23.58495^2)/2) \\
S_{\text{pooled}} &= \text{square root of} ( ((331.07148 + 556.24986)/2) \\
S_{\text{pooled}} &= \text{square root of} ( ((887.32134)/2) \\
S_{\text{pooled}} &= \text{square root of} (443.33037) \\
S_{\text{pooled}} &= 21.05541 \\
\text{d} &= \text{Mean Difference} / S_{\text{pooled}} \\
\text{d} &= -5.75000 / 21.05541 \\
\text{d} &= -0.273 \text{ (absolute value)} = .273 \\
\text{This is a small effect size, e.g. there is little effect on the dependent variable (the mean time to complete simulator views of the two groups) by the independent variable (group assignment) and supports the decision to fail to reject the null hypothesis.}
\end{align*}
\]

Findings statement for independent sample \( t \) –test for the existence of a significant difference in time to complete simulator views in Round Two of assessment between Group I and Group II:

On average, there was no significant difference in Round Two time to complete all simulator views for participants in Group I versus participants in Group II; \( t (15) = -.557, \ p = .586 \) at \( \alpha = .05 \)
Appendix I

SPSS Dependent Sample $t$-tests for Within-Group Comparisons, Round One

SPSS version 24.0.0.0 output for dependent sample $t$-tests, for comparison within Groups in Round One of the experiment. Analysis using standardized assessment scores at baseline and post-treatment Round One.

Dependent sample $t$-test for Group I (initial experimental group), for comparison of baseline assessment scores versus assessment scores after Round One of treatment. Test hypothesis (one-tailed, directional): there was a significant positive difference in before (baseline) and after (Round 1) assessment scores for the participants in Group I. Null test hypothesis: there was no significant positive effect on scores after the first round of the experimental treatment.

**T-Test**

<table>
<thead>
<tr>
<th>Paired Samples Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Pair 1</td>
</tr>
<tr>
<td>BaselineGroup_I</td>
</tr>
</tbody>
</table>

**Paired Samples Test**

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>19.50000</td>
<td>6.18755</td>
</tr>
<tr>
<td></td>
<td>2.18763</td>
<td>2.18763</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.32708</td>
<td>24.67292</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.914</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The critical value of $t$ for $df = 7$ and $\alpha = .05$ is 1.895. With the computed value of $t$ at 8.914 much greater than the critical value, we must reject the null test hypothesis. With the computed with $p$ at .000 ($< 0.001$) much less than our alpha value of .05, we must
reject the null test hypothesis and conclude that there is a statistically significant effect. The decision to reject the null hypothesis is supported by the very large effect size found. Computation of Cohen’s delta ($d$) for effect size for this dependent sample $t$-test:

$$d = \frac{\text{Mean difference}}{\text{S difference}}$$

$$d = \frac{(29.5000 - 10.0000)}{6.18755}$$

$$d = \frac{19.5000}{6.18755}$$

$$d = 3.15$$

This is a huge effect size, (i.e. $> 2.0$) indicating that there is an extremely strong effect by the independent variable (the treatment, in our case the educational unit) on the dependent variable (the scores of participants in Group I in Round One of assessment).

Findings statement for the dependent sample $t$– test for the existence of a significant positive effect by the treatment (the educational unit) on the scores of the participants in Group I: There was a significant positive average difference between baseline and Round One assessment scores for Group I participants ($t_7 = 8.914, p < 0.001$). On average, Round One scores were 19.50 points higher than baseline scores for Group I participants (95% CI [14.32, 24.67]). This was an expected result.

Dependent sample $t$-test for Group II (initial control group), for comparison of baseline assessment scores versus assessment scores after Round One, wherein Group II acted as the control group, receiving only the normally delivered course content and methods. Test hypothesis (one-tailed, directional): there was a significant positive difference in before (baseline) and after (Round 1) assessment scores for the participants in Group II after receiving the normally delivered learning opportunities as the control group (expected result). Null test hypothesis: there was no significant positive effect on Group II scores in Round One.

**T-Test**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pair 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GroupIIRound1</td>
<td>17.667</td>
<td>9</td>
<td>7.19375</td>
<td>2.39792</td>
</tr>
<tr>
<td>GroupIIBaseline</td>
<td>10.222</td>
<td>9</td>
<td>5.38000</td>
<td>1.79333</td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paired Samples Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pair 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GroupIIRound1 -</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GroupIIBaseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Paired Differences</strong></td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.44444</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The critical value of $t$ for $df = 8$ and $\alpha = .05$ is 1.860. With the computed value of $t$ at 4.556 much greater than the critical value, we must reject the null test hypothesis. With the computed with $p$ at .002 much less than our alpha value of .05, we must reject the null test hypothesis and conclude that there is a statistically significant effect. The decision to reject the null hypothesis is supported by the large effect size found. Computation of Cohen’s effect size for this dependent sample $t$-test:

$$d = \frac{\text{Mean difference}}{S \text{ difference}}$$

$$d = \frac{17.6667 - 10.2222}{4.90181}$$

$$d = 1.52$$

This is a very large effect size (i.e. > 1.20) indicating that there is a strong effect by the independent variable (in our case the normally-delivered learning opportunities) on the dependent variable (the Round One scores of participants in Group II). This was expected.

Findings statement for the dependent sample $t$ – test for the existence of a significant positive effect by the normally delivered learning opportunities on the Round One scores of the participants in Group II: There was a significant positive average difference between baseline and Round One assessment scores for Group II participants ($t_8 = 4.556$, $p = 0.002$). On average, Round One scores were 7.44 points higher than baseline scores for Group II participants (95% CI [3.67, 11.21]). This was an expected result.
Appendix J

SPSS Dependent Sample \(t\)-tests for Within-Group Comparisons, Round Two

SPSS version 24.0.0.0 output for dependent sample \(t\)-tests, for within-group comparison of Round Two assessments. Analysis using standardized assessment scores at baseline, post-treatment Round One, and post-treatment Round Two with switching replications experimental design group assignments as initial experimental, then secondary control (Group I), and initial control, then secondary experimental (Group II).

Dependent sample \(t\)-test for Group I (initial experimental group, switching replications control group in Round Two), baseline assessment scores versus scores after Round Two of treatment. Test hypothesis (one-tailed, directional): there was a significant positive difference in before (baseline) and after (Round Two) assessment scores for the participants in Group I after performing the educational unit in Round One and the normally delivered learning opportunities as the secondary control group in Round Two (expected result). Null test hypothesis: there was no significant positive effect on Round Two scores for Group I compared to baseline.

T-Test

Paired Samples Statistics

<table>
<thead>
<tr>
<th>Pair</th>
<th>Round2Group_I</th>
<th>BaselineGroup_I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Deviation</td>
</tr>
<tr>
<td>Pair 1</td>
<td>25.0000</td>
<td>5.01427</td>
</tr>
<tr>
<td></td>
<td>10.0000</td>
<td>3.70328</td>
</tr>
</tbody>
</table>

Paired Samples Test

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15.0000</td>
<td>4.10575</td>
<td>1.45160</td>
<td>Lower 11.56751, Upper 18.43249</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td></td>
<td>10.333</td>
<td>df 7</td>
</tr>
</tbody>
</table>
The critical value of $t$ for $df = 7$ and $\alpha = .05$ is 1.895. With the computed value of $t$ at 10.333 much greater than the critical value, we must reject the null test hypothesis. With the computed with $p$ at .000 ($< 0.001$) much less than our alpha value of .05, we must reject the null test hypothesis and conclude that there is a statistically significant effect. The decision to reject the null hypothesis is supported by the very large effect size found. Computation of Cohen’s delta, effect size for this dependent sample $t$-test:

$$d = \frac{\text{Mean difference}}{S \text{ difference}}$$

$$d = \frac{15.00}{4.10575}$$

$$d = 3.65$$

This is a huge effect size, (i.e. $> 2.0$) indicating that there is an extremely strong effect by the independent variable (the treatment, in this case the long-term effects of the educational unit combined with the normally-delivered course content) on the dependent variable (the scores in Round Two of participants in Group I).

Findings statement for the dependent sample $t$ – test for the existence of a significant positive effect on the Round Two scores compared to baseline scores for the participants in Group I, by the long-term effects of the educational unit combined with the normally-delivered course content: There was a significant positive average difference between baseline and Round Two assessment scores for Group I participants ($t_7 = 10.333, p < 0.001$). On average, Round Two scores were 15.00 points higher than baseline scores for Group I participants (95% CI [11.57, 18.43]). This was an expected result.

Dependent sample $t$-test for Group I (initial experimental group, switching replications control group in Round Two), comparing Round Two assessment scores versus Round One scores. In Round Two Group I acted as the control group, receiving only the normally delivered course content and methods, but still ostensibly enjoying the long-term effects of the educational unit. Test hypothesis (one-tailed, directional): there was a significant positive difference in before (Round One) and after (Round Two) assessment scores for the participants in Group I after receiving the treatment (the educational unit) in Round One and the normally delivered learning opportunities as the control group in Round Two (expected result). Null test hypothesis: there was no significant positive effect on Group I scores compared to baseline.

### T-Test

<table>
<thead>
<tr>
<th>Paired Samples Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
</tbody>
</table>
Paired Samples Test

<table>
<thead>
<tr>
<th>Pair 1</th>
<th>Round2Group_I</th>
<th>25.0000</th>
<th>8</th>
<th>5.01427</th>
<th>1.77281</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round1Group_I</td>
<td>29.5000</td>
<td>8</td>
<td>6.84523</td>
<td>2.42015</td>
<td></td>
</tr>
</tbody>
</table>

Computation of Cohen’s delta, effect size for this dependent sample $t$-test:

$$d = \frac{\text{Mean difference}}{S \text{ difference}}$$

$$d = (29.5000 - 25.0000) / 5.09902$$

$$d = 4.50000 / 5.09902$$

$$d = 0.88$$

This is a large effect size, (i.e. $> 0.8$) indicating that there is a strong effect by the independent variable (the long-term effects of the educational unit combined with the normally-delivered course content) on the dependent variable (the scores in Round Two of participants in Group I).

With the computed value of $t = -2.496$, and the critical value of $t = + or - 2.365$, we must reject the null, and fail to support the test hypothesis. With the computed value of $p$ at .0205 (.041 divided by 2 for a one-tailed hypothesis) being much less than our alpha value of .05, we must reject the null test hypothesis and conclude that there is a statistically significant effect. However, paying attention to the direction of the means we see that it is not a positive effect, but a negative one, and thus cannot support the test hypothesis.

Findings statement for the dependent sample $t$ – test for Round Two scores compared to Round One scores for the participants in Group I: There was a significant negative average difference between Round One and Round Two assessment scores for Group I.
participants ($t_7 = -2.496$, $p = 0.0205$). On average, Round Two scores were -4.50 points lower than Round One scores for Group I participants (95% CI [-8.76, -2.37]). A large effect size was seen of $d = 0.88$. This was not an expected result.

Dependent sample $t$-test for Group II (initial control group, switching replications experimental group in Round Two), baseline scores versus scores after Round Two. Test hypothesis (one-tailed, directional): there was a significant positive difference in before (baseline) versus after (Round Two) assessment scores for the participants in Group II after performing the educational unit in Round Two (expected result). Null test hypothesis: there was no significant positive effect on Round Two scores for Group II compared to baseline.

**T-Test**

<table>
<thead>
<tr>
<th>Paired Samples Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Pair 1</strong></td>
</tr>
<tr>
<td>GroupIIRound2</td>
</tr>
<tr>
<td>GroupIIBaseline</td>
</tr>
</tbody>
</table>

**Paired Samples Test**

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Upper</td>
</tr>
<tr>
<td>t</td>
<td>11.385</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The critical value of $t$ for $df = 8$ and $\alpha = .05$ is 1.860. With the computed value of $t$ at 11.385 much greater than the critical value, we must reject the null test hypothesis. With the computed with $p$ at .000 much less than our alpha value of .05, we must reject the null test hypothesis and conclude that there is a statistically significant effect. The decision to
reject the null hypothesis is supported by the very large effect size found. Computation of Cohen’s effect size for this dependent sample t-test:

\[ d = \frac{\text{Mean difference}}{\text{S.difference}} \]
\[ d = \frac{(29.3333 - 10.2222)}{5.03598} \]
\[ d = 3.79 \]

This is a huge effect size (i.e. > 2.0) indicating that there is an extremely strong effect by the independent variable (in this case the educational unit) on the dependent variable (the Round Two scores of participants in Group II).

Findings statement for the dependent sample t-test for the existence of a significant positive effect by the educational unit on the Round Two scores of the participants in Group II: There was a significant positive average difference between baseline scores and Round Two assessment scores for Group II participants (\( t_8 = 11.385, p = 0.00 \)). On average, Round Two scores were 19.11 points higher than baseline scores for Group II participants (95% CI [15.24, 22.98]). A huge effect size was seen of \( d = 3.79 \). This was an expected result.

Dependent sample t-test for Group II (initial control group, switching replications experimental group in Round Two), comparing Round One scores versus Round Two scores, after treatment. Test hypothesis (one-tailed, directional): there was a significant positive difference in before (Round One) versus after (Round Two) assessment scores for the participants in Group II after performing the educational unit in Round Two (expected result). Null test hypothesis: there was no significant positive effect on Round Two scores for Group II compared to Round One scores.

**T-Test**

**Paired Samples Statistics**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair 1</td>
<td>GroupIIRound2</td>
<td>29.3333</td>
<td>9</td>
<td>5.14782</td>
</tr>
<tr>
<td></td>
<td>GroupIIRound1</td>
<td>17.6667</td>
<td>9</td>
<td>7.19375</td>
</tr>
</tbody>
</table>

**Paired Samples Test**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired Differences</td>
<td>11.6667</td>
</tr>
</tbody>
</table>
The critical value of $t$ for $df = 8$ and $\alpha = .05$ is 1.860. With the computed value of $t$ at 4.730 greater than the critical value, we must reject the null test hypothesis. With the computed $p$ at .0005 (.001 divided by 2 for a one-tailed hypothesis) being much less than our alpha value of .05, we must reject the null test hypothesis and conclude that there is a statistically significant effect. The decision to reject the null hypothesis is supported by the large effect size found. Computation of Cohen’s effect size for this dependent sample $t$-test:

$$d = \frac{\text{Mean difference}}{S \text{ difference}}$$

$$d = \frac{(29.3333 - 17.6667)}{7.39932}$$

$$d = 11.6667 / 7.39932$$

$$d = 1.58$$

This is a very large effect size (i.e. > 1.20) indicating that there is a very strong effect by the independent variable (in this case the educational unit in Round Two) on the dependent variable (the Round Two scores of participants in Group II).

Findings statement for the dependent sample $t$ – test for the existence of a significant positive effect by the educational unit on the Round Two scores of the participants in Group II: There was a significant positive average difference between Round One and Round Two assessment scores for Group II participants ($t_8 = 4.730, p < 0.001$). A very large effect size of $d = 1.58$ was seen. On average, Round Two scores were 11.67 points higher than Round One scores for Group II participants (95% CI [5.97, 17.35]). This was an expected result.
Appendix K

SPSS ANCOVA Analysis of Pre-test/Posttest Scores for Group I and II

SPSS version 24.0.0.0 output for ANCOVA analysis of pretest/posttest scores for Group I and Group II. ANCOVA tests for a statistically significant difference in a dependent variable (post-test scores) between the levels of an independent variable (the experimental group) after controlling for a covariate (the pre-test scores). In our case we will use ANCOVA to test for a statistically significant difference in post-test scores (Round One, the dependent variable) between the levels of the control vs. experimental group (Group I versus Group II, the independent variable) after controlling for the pre-test scores (Baseline, the covariate).

Test hypothesis statement (one-tailed, directional): After controlling for the pre-test scores (Baseline) there was a statistically significant positive difference between Group I and Group II post-test (Round One) scores.

Null test hypothesis: After controlling for pre-test there was no significant difference in post-test (Round One) scores based upon Group level.

There are two assumptions for running ANCOVA. The first assumption is that the pre-test scores cannot be statistically significantly different between the two Groups. This requirement is tested by running ANOVA for the Groups on the pre-test scores:

Univariate Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>8.000</td>
<td>1</td>
<td>8.000</td>
<td>.048</td>
<td>.830</td>
</tr>
<tr>
<td>Intercept</td>
<td>14677.556</td>
<td>1</td>
<td>14677.556</td>
<td>87.547</td>
<td>.000</td>
</tr>
<tr>
<td>Group</td>
<td>8.000</td>
<td>1</td>
<td>8.000</td>
<td>.048</td>
<td>.830</td>
</tr>
</tbody>
</table>
The computed $p$ value for the between-subjects effects for the Groups is shown for the entry for Group as .830, which is non-significant at our $\alpha$ value $= .05$. This satisfies the first assumption that there is no significant variation in pre-test scores between Groups.

Secondly, the assumption of homogeneity of regression is tested for between-subjects effects between Group and pre-test scores (baseline scores) with the post-test scores (Round One) as the dependent variable:

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>7127.911 a</td>
<td>3</td>
<td>2375.970</td>
<td>8.407</td>
<td>.002</td>
</tr>
<tr>
<td>Intercept</td>
<td>4010.891</td>
<td>1</td>
<td>4010.891</td>
<td>14.191</td>
<td>.002</td>
</tr>
<tr>
<td>Group</td>
<td>910.724</td>
<td>1</td>
<td>910.724</td>
<td>3.222</td>
<td>.096</td>
</tr>
<tr>
<td>BaselineScores</td>
<td>1770.103</td>
<td>1</td>
<td>1770.103</td>
<td>6.263</td>
<td>.026</td>
</tr>
<tr>
<td>Group * BaselineScores</td>
<td>18.590</td>
<td>1</td>
<td>18.590</td>
<td>.066</td>
<td>.802</td>
</tr>
<tr>
<td>Error</td>
<td>3674.207</td>
<td>13</td>
<td>282.631</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>85781.000</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10802.118</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. $R^2 = .660$ (Adjusted $R^2 = .581$)

The computed $p$ value for the between-subjects effects for the Groups and the pre-test (Baseline) scores is shown for the entry for Group*BaselineScores as .802, which is non-significant at our $\alpha$ value $= .05$ and thus meeting the assumption for homogeneity of regression.
With both these assumptions met, the ANCOVA analysis can now be run:

### Between-Subjects Factors

<table>
<thead>
<tr>
<th>Value Label</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I = 1 treatment, Group II = 2 control</td>
<td>8</td>
</tr>
<tr>
<td>II = 2 control</td>
<td>9</td>
</tr>
</tbody>
</table>

### Descriptive Statistics

**Dependent Variable: Round One Scores both Groups (post-test)**

<table>
<thead>
<tr>
<th>Group I = 1 treatment, Group II = 2 control</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>84.2500</td>
<td>19.53568</td>
<td>8</td>
</tr>
<tr>
<td>Group II</td>
<td>50.5556</td>
<td>20.37837</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>66.4118</td>
<td>25.98331</td>
<td>17</td>
</tr>
</tbody>
</table>

**Levene's Test of Equality of Error Variances**

<table>
<thead>
<tr>
<th>F</th>
<th>df1</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>.065</td>
<td>1</td>
<td>15</td>
<td>.802</td>
</tr>
</tbody>
</table>

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + BaselineScores + Group

### Tests of Between-Subjects Effects

**Dependent Variable: Round One Scores both Groups (post-test)**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>7109.321*</td>
<td>2</td>
<td>3554.660</td>
<td>13.476</td>
<td>.001</td>
<td>.658</td>
</tr>
<tr>
<td>Intercept</td>
<td>4409.256</td>
<td>1</td>
<td>4409.256</td>
<td>16.716</td>
<td>.001</td>
<td>.544</td>
</tr>
</tbody>
</table>
Baseline Scores

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Scores</td>
<td>2300.925</td>
<td>8.723</td>
<td>.010 .384</td>
</tr>
<tr>
<td>Group</td>
<td>4998.026</td>
<td>18.948</td>
<td>.001 .575</td>
</tr>
<tr>
<td>Error</td>
<td>3692.797</td>
<td>263.771</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>85781.000</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>10802.118</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

a. R Squared = .658 (Adjusted R Squared = .609)

Estimated Marginal Means

1. Group I = 1 treatment, Group II = 2 control

<table>
<thead>
<tr>
<th>Group I = 1 treatment, Group II = 2 control</th>
<th>Mean</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>84.606a</td>
<td>5.743</td>
<td>72.288 96.924</td>
</tr>
<tr>
<td>Group II</td>
<td>50.239a</td>
<td>5.415</td>
<td>38.626 61.853</td>
</tr>
</tbody>
</table>

a. Covariates appearing in the model are evaluated at the following values: Baseline Scores both Groups (pre-test) = 28.8824.

2. Grand Mean

<table>
<thead>
<tr>
<th>Mean</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.423a</td>
<td>3.946</td>
<td>58.960 75.886</td>
</tr>
</tbody>
</table>

a. Covariates appearing in the model are evaluated at the following values: Baseline Scores both Groups (pre-test) = 28.8824.

Levene’s Test for Equality of Error Variances: with our alpha level at .05, the computed value of \( p \) for Levene’s is .802 (> .05), which is not significant, therefore equal error variances are assumed; that is, we fail to reject the null hypothesis of equal error variances between the two groups. This means that there is no significant difference in error variance between Groups.
The computed $p$ value for between-subjects effects for Group is .001 which is statistically significant at our value of .05 so we will decide to reject the null test hypothesis.

The Partial Eta Squared value for Group of .575 indicates that variance in membership from one Group to the other accounts for 57.5% of the variance in the dependent variable, the post-test (Round One) assessment scores. This supports our decision to reject the null test hypothesis.

We consider the direction of the means with our mean values of 84.606 for Group I and 50.239 for Group II. Therefore, we have a finding that there are higher post-test scores for Group I compared to the scores for Group II in Round One of assessment. This supports the directional test hypotheses and the decision to reject the null test hypothesis.

Findings Statement for the ANCOVA test: When controlling for the pre-test scores (baseline assessment scores) there is a statistically significant positive difference in post-test scores (Round One) with higher Group I scores versus the scores in Group II.
Appendix L

SPSS Bivariate Correlational Analysis for Simulator and Clinical Activities

SPSS version 24.0.0.0 output for bivariate correlational analysis of predictive/criterion relationships between Simulator activities and Clinical performance.

A mild negative correlation is found (as expected) between the simulator-based score for obtaining views and structural anatomy identification versus time to complete obtaining all views with the simulator, in Round Two.

### Correlations

<table>
<thead>
<tr>
<th></th>
<th>Time to Complete, Simulator Views</th>
<th>Round 2 Scores, Views &amp; SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to Complete, Simulator Views Pearson Correlation</td>
<td>1</td>
<td>-.183</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.483</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
<tr>
<td>Round 2 Scores, Views &amp; SA Pearson Correlation</td>
<td>-.183</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.483</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
</tbody>
</table>
A weak negative correlation is found between total time to complete all views on the simulator and total time to complete the same views in the clinical setting, an unexpected result.

### Correlations

<table>
<thead>
<tr>
<th></th>
<th>Total Time Simulator Views</th>
<th>Total Time, Clinical Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Time Simulator Views</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
<tr>
<td>Total Time, Clinical Views</td>
<td>Pearson Correlation</td>
<td>-.022</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.934</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
</tbody>
</table>
A mild positive correlation is found between the simulator combined score and the clinical practicum score.

### Correlations

<table>
<thead>
<tr>
<th></th>
<th>Practicum Score</th>
<th>Simulator Combined Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practicum Score</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.478</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Simulator Combined Score</td>
<td>Pearson Correlation</td>
<td>.185</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.478</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>
Next, a very weak almost non-existent positive correlation is found between time to complete obtaining all views with the simulator, versus the clinical combined score (time, views, anatomy) in the clinical setting.

### Correlations

<table>
<thead>
<tr>
<th></th>
<th>Time to Complete, Simulator Views</th>
<th>Clinical Combine Score, Time, Views, &amp; SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to Complete, Simulator Views</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td>Views</td>
<td>Sig. (2-tailed)</td>
<td>926</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
<tr>
<td>Clinical Combine Score, Time, Views, &amp; SA</td>
<td>Pearson Correlation</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.926</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
</tbody>
</table>
Factor Analysis performed to combine data from the various simulator performance factors into one, through the principal component analysis method. This yields a single determinant factor for use as a predictor of criterion outcomes (in our case, clinical performance) and preserves the degrees of freedom versus using the various factors individually.

**Factor Analysis**

**Correlation Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Avg Time to Complete Sim Views, Round 2</th>
<th>Time to Complete, Clinical Views</th>
<th>Round 2 Scores, Views &amp; SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg Time to Complete Sim Views, Round 2</td>
<td>1.000</td>
<td>.024</td>
<td>-.166</td>
</tr>
<tr>
<td>Time to Complete, Clinical Views</td>
<td>.024</td>
<td>1.000</td>
<td>-.038</td>
</tr>
<tr>
<td>Round 2 Scores, Views &amp; SA</td>
<td>-.166</td>
<td>-.038</td>
<td>1.000</td>
</tr>
</tbody>
</table>
### Communalities

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg Time to Complete Sim Views, Round 2</td>
<td>1.000</td>
<td>.546</td>
</tr>
<tr>
<td>Time to Complete, Clinical Views</td>
<td>1.000</td>
<td>.068</td>
</tr>
<tr>
<td>Round 2 Scores, Views &amp; SA</td>
<td>1.000</td>
<td>.563</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

### Total Variance Explained

<table>
<thead>
<tr>
<th>Component</th>
<th>Total</th>
<th>% of Variance</th>
<th>Cumulative %</th>
<th>Extraction Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.177</td>
<td>39.235</td>
<td>39.235</td>
<td>1.177</td>
</tr>
<tr>
<td>2</td>
<td>.990</td>
<td>32.995</td>
<td>72.230</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.833</td>
<td>27.770</td>
<td>100.000</td>
<td></td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

### Component Matrix*

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg Time to Complete Sim Views, Round 2</td>
<td>.739</td>
</tr>
<tr>
<td>Time to Complete, Clinical Views</td>
<td>.261</td>
</tr>
<tr>
<td>Round 2 Scores, Views &amp; SA</td>
<td>-.750</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Using the extracted component, a moderate positive correlation is found between the Simulator Performance Factor as a predictor of clinical performance as measured by the clinical practicum score.
Correlations

Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practicum Score</td>
<td>92.4706</td>
<td>3.64207</td>
<td>17</td>
</tr>
<tr>
<td>Simulator Performance Factor (component)</td>
<td>.0000000</td>
<td>1.00000000</td>
<td>17</td>
</tr>
</tbody>
</table>

Correlations

<table>
<thead>
<tr>
<th></th>
<th>Practicum Score</th>
<th>Component Sim Performance Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practicum Score</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.227</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
<tr>
<td>Component Sim Performance Factor</td>
<td>Pearson Correlation</td>
<td>.227</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.380</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>17</td>
</tr>
</tbody>
</table>

![Graph showing correlation between Practicum Score and Component Sim Performance Factor]
Next, using the extracted component, a very weak positive correlation is found between overall Simulator Performance Factor as a predictor of clinical performance as measured by the clinical performance score.

**Correlations**

### Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Score (BTTEET)</td>
<td>26.0000</td>
<td>4.74342</td>
<td>17</td>
</tr>
<tr>
<td>Component Sim Performance Factor</td>
<td>.0000000</td>
<td>1.000000000</td>
<td>17</td>
</tr>
</tbody>
</table>

### Correlations

<table>
<thead>
<tr>
<th></th>
<th>Clinical Score (BTTEET)</th>
<th>Component Sim Performance Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Score (BTTEET)</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.028</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>Component Sim Performance Factor</td>
<td>Pearson Correlation</td>
<td>.028</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.915</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>17</td>
</tr>
</tbody>
</table>
A Factor Analysis was then performed to combine data from the various clinical data factors into one, through the principal component analysis method.

**Factor Analysis**

**Correlation Matrix**

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Time to Complete, Clinical Views</th>
<th>Clinical Score (BTTEET)</th>
<th>Practicum Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>.000</td>
<td>1.000</td>
<td>-.045</td>
</tr>
<tr>
<td>Time to Complete, Clinical Views</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Score (BTTEET)</td>
<td>-.045</td>
<td>1.000</td>
<td>.868</td>
</tr>
<tr>
<td>Practicum Score</td>
<td>.035</td>
<td>.868</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Communalities**

<table>
<thead>
<tr>
<th>Initial Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to Complete, Clinical Views</td>
</tr>
</tbody>
</table>
Clinical Score (BTTEET)  1.000  .935  
Practicum Score  1.000  .934  

Extraction Method: Principal Component Analysis.

<table>
<thead>
<tr>
<th>Component Matrix*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Component</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Time to Complete, Clinical Views</td>
</tr>
<tr>
<td>Clinical Score (BTTEET)</td>
</tr>
<tr>
<td>Practicum Score</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
a. 1 components extracted.

Finally, using the second, newly extracted component, another mild positive correlation is found between the Simulator Performance Factor and as a predictor of clinical performance as measured by the Clinical Performance Factor.

<table>
<thead>
<tr>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component Clinical Performance Factor</td>
</tr>
<tr>
<td>Component Clinical Performance Factor</td>
</tr>
<tr>
<td>Performance Factor</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Component Sim</td>
</tr>
<tr>
<td>Performance Factor Pearson Correlation</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>N</td>
</tr>
</tbody>
</table>
Appendix M

Photograph of the UI of an Actual Ultrasound Imaging Machine

Photograph of the UI for an actual ultrasound imaging machine as used in the clinical lab setting of the learner participants involved.
Appendix N

Photographs of the UI for the Simulator

Photograph of the physical UI for the simulator used in the setting of this study.
Photograph of the on-screen virtual UI controls for the simulator used. Below is a close-up photograph of the on-screen virtual UI controls for the simulator used.
References


Association for Medical Education in Europe (2015). Retrieved from https://www.amee.org/what-is-amee


Damewood, S., Jeanmonod, D., & Cadigan, B. (2011). Comparison of a multimedia simulator to a human model for teaching FAST exam image interpretation and image acquisition. *Academic Emergency Medicine, 18*, 413-419.


