University of Rhode Island

DigitalCommons@URI

2020

STRUCTURED, INDIVIDUALIZED, VIDEO-BASED COACHING TO IMPROVE TECHNICAL SKILL ACQUISITION

Kenneth A. Lynch Jr.
University of Rhode Island, kenneth_lynch@brown.edu

STRUCTURED, INDIVIDUALIZED, VIDEO-BASED COACHING

TO IMPROVE TECHNICAL SKILL ACQUISITION

BY

KENNETH A LYNCH, JR

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

EDUCATION

UNIVERSITY OF RHODE ISLAND

AND

RHODE ISLAND COLLEGE

2020

DOCTOR OF PHILOSOPHY DISSERTATION

OF

KENNETH A. LYNCH, JR.

APPROVED:

Dissertation Committee:

Major Professor: Dr. Kathy Peno

Dr. Minsuk Shim

Dr. Kara Misto

Dr. Joanne Costello

Dr. Stefan Battle

Dr. Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2020

Abstract

In the current era of a reduced hourly workweek and competency-based curriculum, it is becoming increasingly clear that the cognitive apprenticeship model of training, which dominated the surgical discipline for well over a century, may no longer be the single best strategy to train the surgeons of today. There needs to be a multifaceted approach to surgical residency training, especially with regards to improving technical skill acquisition. Video-based-coaching is quickly being adapted for use in many training programs due to its low cost and broad range of applications including self-assessment, operative evaluations, and bench model training. The challenge for most surgical residency programs is that there is very little discussion regarding the best way to train faculty as coaches, which particular coaching framework should be used, or how best to incorporate coaching into a training curriculum. The purpose of this research is to address these questions by comparing the implementation of a curriculum using structured, individualized, video-based coaching against self-assessment video analysis only as a way to improve technical skill acquisition in surgical residents participating in a vascular anastomosis simulation workshop.

Acknowledgements

I would like to take this opportunity to thank my family, friends, and colleagues for their support throughout this journey. Thank you to the surgical residents who I have the pleasure to interact with and work beside on a daily basis throughout my career. They are the hardest working individuals I know and their drive to improve performance was the inspiration for this research.

I want to take a moment and extend a special thank you to my committee chair, Dr Kathleen Peno, for her ongoing support, constructive feedback, guidance and advice from the moment I inquired about pursuing my PhD in Education. She is everything a student could ask for in a mentor and I am forever grateful for the opportunity she provided me. A special thank you to Dr Shim for opening up a world of statistics that I didn't realize I could ever fully grasp or become comfortable with. I would also like to extend my sincere gratitude to the rest of my committee members. Dr Battle, Dr Costello, and Dr Misto, thank you for dedicating your time and effort to this process. Your encouragement, backing, guidance, and mentorship pushed me to think more broadly about this research and your input was instrumental to my growth as a researcher.

Finally, I could not be where I am today without the sacrifice and support from my wife, Tara, and my two children, Aidan and Kailynn. I am forever indebted to them for their endless encouragement and optimism throughout this process. They continue to inspire me to be a better husband, father, clinician, researcher, and individual.

Table of Contents

List of Figures

List of Graphs

Chapter One: Introduction

If John Dewey was correct in his description of the association between experience and education, then "everything depends upon the *quality* of the experience which is had" (Dewey, 1938, p. 26). The education of surgical residents has undergone radical evolutionary changes over the past two decades, including a change to competency-based education, a mandatory reduction in the hourly resident workweek, and an intense focus on error free-practice. Because of these changes, residency programs have recognized that the apprenticeship model of training, pioneered by William Stewart Halsted in the early 20th century, and based on the "see one, do one, teach one" method, may no longer be the most effective strategy to prepare surgical residents for practice (Gallagher, Jordan-Black, & O'Sullivan, 2012). In an attempt to strengthen surgical training, and ensure we are providing the public with educated, well-trained, technically competent surgeons, it is the *quality* of our residency training programs that surgical educators are currently trying to improve upon.

To attain this goal, graduate medical education has significantly increased its adoption and implementation of adult learning theory into its curriculum in order to improve resident experience, facilitate transfer of learning, and improve knowledge and skill acquisition. As noted by Carter and Gogia (2014), however, much of this implementation has been done without the input of qualified educators, leading to inconsistent results and varying degrees of success amongst programs. The basis of this dissertation research is specifically targeted at the acquisition of technical skill in junior surgical residents. This complex subject matter incorporates a variety of theoretical foundations of adult learning theory including experiential learning, social

constructivism, motor learning theory, distribution of practice, deliberate practice, coaching, a model of skill acquisition, self-assessment, and feedback. The purpose of this dissertation is to explore these theories and concepts, specifically as they relate to surgical training, explain how they may help improve transfer of learning and skill acquisition, and discuss the current limitations that exist with regards to implementing these theories into practice.

Background

Apprenticeship Model of Training

The Oxford English Dictionary defines an apprentice as: "A learner of a craft, bound to serve, and entitled to instruction from, his or her employer for a specified period" (Oxford English Dictionary, 2017). The most common method for training surgeons throughout the ages has, indeed, been through variations in the apprenticeship model. While there was no set standard for these apprenticeships, an individual in the 16th century would begin their journey around the age of twelve or thirteen, and remain in an apprenticeship for five to seven years (Polavarapu, Kulaylat, Sun, & Hamed, 2013). In its most basic form, the student learns the art and science of surgery through the direct observation and imitation of the actions of their mentor in both the clinical and operating room environment (Dunnington, 1996). The main advantage to a natural apprenticeship model, such as this, is that learning is situated in the context of a specific activity, where skills are directly observed. As Lajoie (2010) points out, however, the pitfall to this natural apprenticeship model, is that "cognitive skills are not readily observable and adaptations to such an apprenticeship model are necessary to support education more broadly" (p. 64).

Most apprenticeship models in place today, however, are based on the cognitive apprenticeship model. A cognitive apprenticeship is defined by Collins, Brown, & Newman (1989) as "learning through guided experience on cognitive and metacognitive, rather than physical, skills and processes" (p. 456). This form of apprenticeship revolves around a complex interaction between a mentor and student that is dependent on expert demonstration and careful guidance, especially in the early stages of learning. Building off this cognitive apprenticeship model, and recognizing that learning occurs through social interaction, Lave and Wenger (1991) developed their theory of situated learning and the model of legitimate peripheral participation. This theory attempts to explain the way in which learning occurs through socialization within what they call "communities of practice." More specifically, this model concerns the "relations between newcomers and old-timers, and about activities, identities, artifacts, and communities of knowledge and practice" (p. 285-286).

Understanding our current model of surgical training allows us to better appreciate how, over the course of five years, surgeons move from their first intern year, where they are mostly observers on the periphery of participation, to their fifth and final chief year, during which they become functionally independent and are full participants within the surgical community of practice. This apprenticeship model, both in its natural and cognitive form, has been the mainstay of surgical training since its inception, and ultimately became the education gold standard at every surgical residency program offered throughout the world. How this process occurred, however, becomes more apparent when one takes a closer look at the way surgical training evolved from a trade to a profession over the last few centuries.

Historical Perspective

Europe. The first efforts to improve surgical training date back to Paris, 1210 AD, at the College of de Saint Come (Franzese & Stringer, 2007). It was here that attempts were made to identify academic surgeons, who had acquired training or attended a university, with the distinction of donning a long robe, from the less reputable, or short-robe, barber-surgeons (Bodemer, 1983). Despite this distinction, however, surgeons were far from being considered "equal" to their medical physician counterparts (Loudon, 2000). The medical physicians supposed superior knowledge throughout this era gave them the theoretical authority to oversee the work of surgeons (Cope, 1959). This division between the two disciplines, however, would soon begin to weaken with the rapid growth of the volunteer hospitals in Europe between 1730-1800. This, along with the work of anatomist and surgeon John Hunter, allowed the "hospital surgeon" to emerge as a consultant of high professional rank (Temkin, 1951).

It was this group of surgeons that ultimately led to the formation of the Company of Surgeons in 1745 (Morgan, 1968). The Company of Surgeons recognized the importance of providing formal lectures, along with an apprenticeship, as part of the education and training of future surgeons. According to the Company bylaws, each "Master' surgeon could have up to three apprentices, and each were required to provide seven years of servitude (Wall, 1937). In addition to this, Company of Surgeons helped with the planning and construction of an anatomy theater, close to Newgate prison in West London, in order to facilitate the dissection of deceased criminals for instructive purposes (Robinson, 1984).

A key aspect of the Company's role, however, was to register and oversee a surgeons' apprenticeship, and then formally examine these surgeons upon their completion (Tyte, 2011). Upon completion of their seven years, examinations were to take place in Latin, and candidates were questioned on all aspects of surgery, which was defined as:

> not only the external and actual practice thereof, but also the internal speculation of the natural causes and remedies of all manner of infirmities or diseases incident to the said practice and profession and of the natures and qualities of all manner of emplasters, ointments, medicaments, baths, waters, drugs, and herbs pertaining thereunto. (Wall, 1937, p. 49)

Due to their advancement of the craft of surgery, the Company of Surgeons was granted a new charter by King George the III on March 22, 1800, bestowing upon them the title of the Royal College of Surgeons (Cope, 1959). Despite this professional achievement however, the division between medicine and surgery continued to persist in Europe. During granting of the charter, Lord Thurlow, a member of the house of lords, commented, "There is no more science in surgery than in butchering" (Rogers & Münsterberg, 1906, p. 306). It wasn't until the 1850's when Germany began the standardization of examinations and licensing, as well as the recognition of formal specialties including surgery, that the process of substantial medical reform began taking place in Europe (Shryock, 1965).

United States. While this transformation of the education and training of surgeons was slowly gaining some momentum in 18th Century Europe, the same cannot be said for the United States. Most medical practitioners in colonial America had little to

no formal education, as they were either self-taught, or came into their craft through informal apprenticeships (Goldowsky,1988; Warren, 1958). Out of the nearly 4000 practitioners in 1776, only around 400 of them had a formal medical degree, and the majority of these came from the Edinburgh school of Medicine in Scotland (Bordley, 1976). Colonial practitioners of medicine, unlike their European colleagues, were expected to be a "jack of all trades," encompassing medicine, surgery, and handing out herbal remedies and medicines of the time (Grillo, 1999). In fact, in 1690, when congregationalist minister Samuel Lee of Massachusetts was asked about the undifferentiated character of medical practice in New England, he wrote: "Practitioners are laureated gratis with the title feather of Doctor. Pothecaries, surgeons & midwives are dignified according to success." (Gill, 1972, p. 19).

Aside from the lack of formal education available for those who wanted to practice medicine and surgery in colonial America, there was another factor limiting its progression. Unlike Europe, where hospitals had flourished for centuries, the only facilities that existed in America were almshouses, built by charities, that were set up to help deal with the old and the ailing poor (Stain, 2015). It wasn't until the first hospital in America, The Pennsylvania Hospital, opened its doors in 1751, that medical, and particularly surgical education, began to take shape on our soil (Williams, 1972). It was in this institution that the current apprenticeship model began to formally develop as "attending Physicians brought with them their students, or apprentices, to follow the practice of the house, to apply dressings, and render other assistance" (Morton & Woodbury, 1895, p. 479). This arrangement ultimately led to the opening of the first medical school in the original 13 colonies in 1765, The Medical College of Philadelphia

(Fee, 2015). Following in the footsteps of the European model of medical schools, particularly Edinburgh, this medical college set strict requirements for admission, which included; education in the liberal arts, mathematics, and natural history; knowledge of Latin and preferably French; and a required apprenticeship of no less than 3 years with a reputable physician (Flexner, 1910).

The 19th century was one of rapid growth for medical and surgical education, with hospitals and medical colleges opening their doors in most major cities. Unfortunately, oversight for these institutions was severely lacking, and it became apparent that the only state requirement for licensure was a medical school diploma (Miller & Weiss, 2008). Sadly, even applicants who were barely literate were granted admission to these medical colleges, as the for-profit institutions competed to provide applicants with "the fastest, easiest, and cheapest education" (Kaufman, 1976, p. 42) as opposed to the best training. It was this culmination of the questionable benefit of the education offered at these colleges, combined with insufficient medical knowledge, improper training, and laxity in state licensure which drove the formation of the American Medical Association (AMA) in 1847. As such, they saw to it that one of their first mandates was to improve medical training in America and assure physician competencies (King, 1983). Because of the incompetence of American medical education and surgical training during this time, students who wanted to further their education often turned to Germany and Austria (Grillo, 1999). One surgeon, in particular, decided to do just that, and forever changed the landscape of surgical training in the United States.

Halstead, the German System, and Churchill. William Stewart Halstead received an undergraduate degree from Yale in 1874, and his MD degree from the

7

College of Physicians and Surgeons of New York in 1877. After practicing as a house physician at Bellevue Hospital in New York City from 1877-1878, Halstead realized he had acquired all of the training available to him in the United States, so he travelled to Europe to further his surgical studies in Vienna and Würzburg from 1878-1880 (Cameron, 1997). It was during this time, Halsted became acquainted with the surgical residency training model in Germany. This particular training model, which was much different than that offered in the United States, was built upon the premise that surgical trainees should receive an increasing amount of responsibility with each advancing year (Hamdorf & Hall, 2000).

It was this "pyramidal" model that Halstead sought to replicate when he joined the staff of the newly built Johns Hopkins Medical Center in Baltimore in 1889. Modeled after what he had seen in Germany, this surgical apprenticeship was based upon the selection of eight interns annually. From this group, four stayed on for only one year, while the other four remained for varying terms. Of the initial eight, only one of them ultimately became the senior house surgeon, with the other three waiting their turn for preferment. The average term for one of these house surgeons was eight years in total (Grillo, 1999). It is important to note that in this model, resident advancement was not guaranteed. Halsted felt that this rigorous training in surgery would help produce academic leaders in both surgical practice and surgical research. He described the purpose of his residency program as follows: "We need a system, and we shall surely have it, which will produce not only surgeons, but surgeons of the highest type, men who will stimulate the first youths of our country to study surgery and to devote their energy and their lives to raising the standard of surgical science" (Halsted, 1904, p.273).

Halsted's pyramidal model greatly improved the structure and standardization of surgical training in the United States, and in 1928 the American Medical Association House of Delegates approved the principals of this model to guide surgical residencies and fellowships (Hamdorf & Hall, 2000). Not everyone, however, bought into this model as they best way to train a surgeon. Edward D. Churchill, who became chief of surgery at Massachusetts General Hospital (MGH) in 1931, took issue with the strict "pyramidal" training system the Halsted model encouraged, particularly its focus on producing a single master-apprentice. Churchill believed in a more broad-based surgical training system, one where "the school concept of a group of masters, in which no single personality dominates the methods and technology of the institution" (Grillo, 2004, p. 949).

Based on this philosophy, Churchill proposed a model of surgical training where he would select six residents a year, for training that would last four to five years. At the end of this time, these residents would be fully trained in general surgery and be capable of independent practice (Helling, 2016). This "rectangular" model of surgical training was put into effect in 1940 at MGH in substitution of the "pyramidal" system popularized by Halsted. According to Grillo (2004), this model of training proved extremely successful because it " (1) succeeded in giving complete training to superior surgeons, (2) it eliminated the human wastefulness of the pyramidal system, and (3) it responded to the nation's surgical needs after World War II" (p. 951). Because of this, the Halsted model of surgical training that had been in place for over 50 years, was ultimately replaced by Churchill's rectangular system, as it better met the needs of both the surgeons and society as a whole.

Present Day Surgical Training. Today, general surgery training in the United States is a five-year program that may be extended if the candidate chooses to pursue research training, which could lengthen this duration (Accreditation Council for Graduate Medical Education, 2019). While the rectangular model of resident selection, in conjunction with an apprenticeship model of training, are still the basis for surgical residency, graduate medical education has undergone significant changes since the start of the 21st century. These changes include a focus on discipline specific core competencies, the institution of a reduced, 80-hour workweek for residents, and an emphasis on patient safety and error free practice. Because of this, the ability to define competent performance in practice has quickly become a major focus for all members of the American Board of Medical Specialties (ABMS). This effort was best highlighted in the 2001 ABMS, Accreditation Council for Graduate Medical Education (ACGME) joint report on surgical competencies, and the criteria needed for maintenance of certification (Ritchie, 2001). The impetus for these national initiatives were, in part, driven by two major events at the end of the 20th century.

The first event was the tragic, unexpected death of a young woman, Libby Zion, in a New York hospital in 1984. Libby was a college freshman with a history of depression who was admitted to the emergency room with symptoms of fever, agitation and disorientation (Lerner, 2006). Six hours after Libby's admission, she had died from a cardiac arrest. Her death resulted from a drug interaction between her home antidepressant and a medication prescribed to her to control her agitation. Her care during those six hours was managed by an emergency room resident and an intern, and no attending physician was called, even as her condition deteriorated. After a long court

battle, the investigation ruled that her death was a direct result of lack of resident oversight and resident fatigue (Patel & Popp, 2014).

The second event was the Institute of Medicine's (IOM) original publication "To Err is human" (Donaldson, Corrigan, & Kohn, 1999). In an effort to address patient safety within health care, data was randomly collected over a fifteen-year period from multiple healthcare institutions in New York, Utah, and Colorado in search of adverse events. The researchers defined adverse events as "an injury that was caused by medical management (rather than the underlying disease) and that prolonged hospitalization, produced disability at the time of discharge, or both" (Brennan et al., 1991). The culmination of this research found health care to be far behind other industries at ensuring basic safety and, based on the data, estimated that as many as 98,000 people die annually in American hospitals due to medical error. The report called for immediate action from all agencies to help reduce this number of deaths by 50% over the following five years.

In response to these two events, the ACGME implemented its resident outcomes project in July 2001, which delineated six specific core competencies of medical training against which all residents would be assessed (Mery, Greenberg, Patel, & Jaik, 2008). These core competencies, which are now fully integrated into all ACGME residency programs, include medical knowledge, patient care, professionalism, interpersonal and communication skills, practice-based learning and systems-based learning (Swing, 2007). In addition to these competencies, and in an effort to address patent safety concerns and resident well-being, the ACGME moved forward with a plan to reduce the hourly workweek for all residency training programs. In July, 2003, the 80-hour residency

workweek officially went into effect (Philibert, Friedmann, & Williams, 2002), and dramatically altered the surgical training landscape.

Statement of the Problem

By 2010, it became apparent that the resident work-hour restriction was creating a new set of problems associated with all residency training. These problems included a growing trend towards a shift-work mentality, programs paying too much attention on duty hours and less focus on resident supervision, as well as the creation of ethical dilemmas for resident providers, such as having to choose between leaving patients to comply with duty hours, or staying to provide care and violate hours (Nasca, Day, & Amis, 2010). The cumulative effect of these work-hour restrictions had even further repercussions for surgical training, however. Despite an upward trend in overall surgical case volume, general surgery residents were experiencing a decrease in their operative case exposure (Watson, Flesher, Ruiz, & Chung, 2010). In another survey of surgical residency program directors, Mattar et al. (2013) reported that, of graduating chief surgical residents, 66% were unable to perform a major procedure unsupervised for more than 30 minutes, and for those performing laparoscopic procedures, 26% could not identify anatomic planes, 56% could not suture, and 30% could not independently perform a laparoscopic cholecystectomy.

Although another work-hour restriction was made in 2011 to limit first year residents to only 16 continuous hours of duty, this was reversed by the ACGME in the 2017 common requirements. This reversal was due to multiple studies showing, in part, a negative impact on resident education (Bolster & Rourke, 2015), an unintended increase in self-reported medical errors (Sen et al., 2013), and a significant decrease in operative

experience for first-year surgical residents (Schwartz et al., 2013). Given this trend in

general surgery training, surgical residents are, themselves, feeling less competent in their

technical skills upon graduation. Coleman, Esposito, Rozycki, & Feliciano (2013)

reported that nearly 40% of residents lacked confidence in their skills after five years of

training and Fronza, Prystowsky, Darosa, & Fryer (2012) reported that graduating

residents are feeling less competent in their ability to perform many "core" general

surgical operations upon graduation. This was again echoed in a more recent review of

14 general surgery programs involving 872 faculty, 511 categorical residents, and over

10,000 procedures. In this study, George et al., (2017) found that US general surgery

residents are still not universally ready to independently perform common "core"

procedures upon completion of residency training. It appears that in our current training

model, an unintended consequence of the policies put in place to improve patient safety,

resident lifestyle and reduce medical errors have unintentionally affected the technical

competency of our residents graduating from surgical training programs.

Given this trend in present day surgical residency, faculty educators continue to

look for novel ways to improve surgical training and technical skill acquisition, both

inside and outside the operating room. Because of its applicability to a wide range of

specialties, simulation training has continued to gain traction over the last twenty years as

a means of improving residents' technical skills. Simulation can be defined as "a

technique to replace or amplify real experiences with guided experiences, often

immersive in nature, that evoke or replicate substantial aspects of the real world in a fully

interactive manner" (Gaba, 2004). The use of simulation training in surgery dates back to

ancient India in 600 BC, where leaf and clay models were used to simulate the very first

recorded operation, the forehead flap nasal reconstruction (Limberg, 1984). Since that time, surgeons have primarily looked to using cadavers and/or animal models to simulate surgery, but the downsides of cost, availability, anatomical variation, and ethical considerations can be prohibitive for many institutions (Rosen, Long, McGrath, & Greer, 2009). Because of these limitations, the use of inanimate bench models to teach technical skills were quickly adopted by most surgical training programs due to their low cost, portability, and ability for repeated use.

Specialty specific bench model workshops were developed nearly forty years ago in the United Kingdom by the Royal College of Surgeons (Bevan, 1981). It is here that one of the first vascular anastomosis bench model simulation classes was held, which employed the use of a newly created Buckston Browne Arterial Jig, developed by vascular surgeon Dr. Roger Greenhalgh (Greenhalgh & Flack, 1981). This jig allowed easy demonstration and practice of the proper anastomosis suturing technique of an open aneurysm repair under direct supervision.

After the initiation of these workshops and the development of the Buckston Browne Arterial Jig in the 1980's, simulation workshops became much more common in surgical training programs. These early anastomosis bench models, however, were conducted largely without an emphasis on objective measurements of the participants. It was assumed, in these early classes, that the act of attending and participating in these workshops would ultimately ensure a resident's technical skill acquisition (Atkins, Kalu, Lannon, Green, & Butler, 2005). More recently, however, various models for evaluating a resident's technical skills associated with participation in vascular anastomosis bench

models have been well described and validated in the literature (Datta et al., 2002; S. Schwartz et al., 2014; Wilasrusmee, Phromsopha, Lertsitichai, & Kittur, 2007).

Research has also shown a transferability of skills and a positive correlation between simulation training and technical competency in the operating room and clinical setting. Using a vascular anastomosis simulation workshop, Wilasrusmee, Lertsithichai, & Kittur (2007) found that time to complete anastomosis and grade of anastomotic leak was predictive of technical competency in the operating room. These positive results have also been shown in other specialty simulation workshops. In a review of twenty randomized control trials in laparoscopic surgery, Vanderbilt et al., (2015) found evidence of improved clinical performance in surgeons who underwent simulation-based training. In another systematic review of twenty-seven randomized controlled trials and seven comparative studies encompassing both laparoscopy and endoscopy, Dawe et al., (2014) concluded that "these studies provided strong evidence that participants who reached proficiency in simulation-based training performed better in the patient-based setting than their counterparts who did not have simulation-based training" (p. 1063). Perhaps even more importantly, however, is the evidence suggesting an association between simulation training and improved patient outcomes (Barsuk et al., 2018; Brydges, Hatala, Zendejas, Erwin, & Cook, 2015). These outcomes have led surgical training programs to ultimately place a much greater emphasis on using simulation bench model workshops to improve junior resident's performance on competency-based procedural skills.

Advances in technology have also made alternative surgical training modalities, such as virtual reality and robotic trainers, very attractive to program directors and

educators, but their implementation remains challenging and they continue to remain cost prohibitive for most residency programs (Nagendran, Gurusamy, Aggarwal, Loizidou, & Davidson, 2013). Due to its low cost and broad range of applications including self-assessment, evaluations, and use in bench model simulation, the use of video recording, with playback and analysis, has quickly found its way into residency training programs over the last decade. Building on this concept, some programs have now begun to incorporate video-based coaching as a method to help improve technical skill acquisition, but its use has not been standardized, leading to varied implementation and inconsistent results (Min, Morales, Orgill, Smink, & Yule, 2015).

What makes these efforts more complicated, however, is that despite the major policy initiatives that began at the turn of this century, surgical faculty and educators struggle to agree upon a validated, reliable way to assess technical competency in surgical training. Although the concept of assessing residents based on clinical-based competencies was initiated in 2001, it wasn't until the ACGME Milestone Project in 2014 that "technical skill" was added to the list of required competencies for surgical residents (Cogbill, Malangoni, Potts, & Valentine, 2014). Another reason assessing technical skill has proven difficult is because there continues to be a lack of consensus on a formalized definition of technical competency. Szasz, Louridas, Harris, Aggarwal, & Grantcharov (2015) note that one explanation for this is because the literature tends to use the terms competence and proficiency interchangeably. This is an incredibly important distinction, and one that this research intends to address using a model of skill acquisition.

Purpose of the Study

The purpose of this research is to employ a validated technical skills checklist to compare structured, individualized, video-based coaching against self-assessment video analysis as a way to improve technical skill acquisition in surgical residents participating in a vascular anastomosis simulation workshop. The hypothesis for this research project is structured, individualized, video-based coaching improves technical skill acquisition to a greater extent than self-assessment video analysis.

Research Questions

In order to evaluate the effect, if any, that structured, individualized, video-based coaching may have on technical skill acquisition, the following research questions were used to help guide this study:

RQ1A: Based on expert evaluator scoring using a validated technical skills checklist, to what extent do residents who participate in structured, individualized, video-based coaching attain technical skill, compared to residents in a self-assessment, video analysis only group?

RQ1B: Using the validated technical skills checklist, to what extent do resident's self-assessment scoring of their own technical skill differ between those who participate in the structured, individualized, video-based coaching group, compared to those in the self-assessment, video analysis only group?

RQ2: Based on expert evaluator scoring using the validated technical skills checklist, do residents who participate in structured, individualized, video-based coaching attain technical skill at a different rate than residents in the self-assessment video analysis only group?

RQ3: Based on the validated technical skills checklist, to what extent do resident self-assessment scores of their technical skill correlate with expert evaluator scores?

RQ4: To what extent do the current posttest scores on time to fashion anastomosis and leak rate compare with the prior three years of resident vascular anastomosis scores?

Significance of the Study

In the current era of surgical residency training, faculty and educators need to structure curriculum in a way that efficiently promotes the transfer of learning and improves technical skill acquisition. Coaching, combined with playback video analysis, has been increasingly employed as a way to improve technical skill acquisition (Bonrath, Dedy, Gordon, & Grantcharov, 2015; Soucisse et al., 2017) but how this compares to self-assessment video review has yet to be explored in the surgical literature. Until now, our own local surgical residency program has not included any of this type of technology in our simulation labs. With regards to our vascular anastomosis simulation lab, specifically, technical skills have been assessed using only two primary variables; time to perform anastomosis and leak rate.

I expect this research to demonstrate that it is feasible to train surgical faculty on how to adapt a structured coaching format into a surgical training curriculum in order to promote transfer of learning and improve technical skill acquisition. The findings from this research will add to the current literature on the use of video analysis for surgical assessment purposes and help further define the role of self-assessment video analysis as it pertains to technical skill acquisition. The goal of this research is to adapt what I learn from this simulation model and transition this to the operating room environment,

specifically in training residents how to perform common core general surgical procedures such as laparoscopic cholecystectomy, and laparoscopic appendectomy.

Definition of Key Terms

To aid the reader, this section includes definitions of common terms that are used throughout the study.

Andragogy: "The art and science of helping adults learn" (Knowles, 1980, p.43).

Apprentice: "A learner of a craft, bound to serve, and entitled to instruction from, his or her employer for a specified period" (Oxford English Dictionary, 2017).

Coaching: "The art of creating an environment, through conversation and a way of being, that facilitates the process by which a person can move toward desired goals in a fulfilling manner" (Timothy Gallwey, 2002, p.177).

Cognitive apprenticeship: "Learning through guided experience on cognitive and metacognitive, rather than physical, skills and processes" (Collins, Brown, & Newman, 1989, p. 456).

Consolidation theory: Memories remain in a state that is vulnerable to disruption immediately after learning a new task, and take time to become fixed (or consolidated) (Lechner, Squire, & Byrne, 1999).

Deliberate practice: "Individualized training activities especially designed by a coach or teacher to improve specific aspects of an individual's performance through repetition and successive refinement" (Ericsson & Lehmann, 1996, p. 278).

Feedback: "Specific information about the comparison between a trainee's performance and a standard, given with the intent to improve the trainee's performance" (Van De Ridder, Stokking, McGaghie, & Ten Cate, 2008).

Goals: "A cognitive image of an ideal stored in memory for comparison to an actual state; a representation of the future that influences the present; a desire (pleasure and satisfaction are expected from goal success); a source of motivation, an incentive to action" (Cochran & Tesser, as cited in Street, 2002, p.100).

Knowledge of Results: "Information, and how the subject will transform and use it will depend on the type and accuracy of knowledge of results and on the kind of motor task" (Adams, 1976, p. 90).

Knowledge of Performance: "Augmented information about the movement pattern the learner has just made" (Schmidt et al., 2018, p. 438).

Learning: "The process whereby knowledge is created through the transformation of experience" (Kolb & Kolb, 2005).

Massed training session: "Those in which individuals practice a task continuously without rest" (Donovan & Radosevich, 1999, p. 795).

Mentor: "A voluntary and active participant in the personal and professional development of the mentee, offering knowledge, experience, guidance, support, and opportunity for advancement" (Strowd & Reynolds, 2013, p. 244).

Reflecting-on-Practice: a way for practitioners to "think back on a project they have undertaken, a situation they have lived through, and they explore the understandings they have brought to the handling of the case" (Schon, 1983, p. 61).

Self-Assessment: ''The involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they have met these criteria and standards'' (Boud, 2013, p. 12).

Simulation: "A technique to replace or amplify real experiences with guided experiences, often immersive in nature, that evoke or replicate substantial aspects of the real world in a fully interactive manner" (Gaba, 2004).

Spaced training session (Distributive practice): "Those in which individuals are given rest intervals within the practice session" (Donovan & Radosevich, 1999, p. 795).

Technical skill: "The ability to use surgical instruments in an effective and efficient manner and includes economy of motion and safe tissue handling" (Mellinger et al., 2017, p. 52).

Transfer of Learning: "The effective application by program participants of what they learned as a result of attending an education or training program" (Caffarella, 2012, p 5212-5213).

Chapter 1 Summary

This chapter presented an overview and discussion on the evolution of the apprenticeship model of surgical training in both Europe and the United States. Landmark events surrounding medical and surgical training at the end of the 20th century were identified in relation to the national policies put in place by the ACGME to ensure patient safety and reduce medical errors through competency based medical and surgical education. The purpose of the study was to employ a validated technical skills checklist to compare structured, individualized, video-based coaching against self-assessment video analysis as a way to improve technical skill acquisition in surgical residents participating in a vascular anastomosis simulation workshop. While a current review of the available literature has shown significance when comparing coaching to standardized training, there has not been any literature comparing coaching to self-assessment when

benchmark videos are provided. The intent of this study is to add to the current literature on the use of video analysis, coaching, and self-assessment in relation to technical skill acquisition in junior surgical residents.

Organization of Study

This research is organized into five chapters. Chapter one introduced readers to the history of the surgical apprenticeship model and the current status of surgical residency training in the United States. The chapter also introduced the reader to the statement of the problem, purpose of the study, and research questions. The significance of the study was also discussed, including definition of key terms, and organizational structure of this research.

Chapter two will focus on an in-depth review of the literature, including key concepts and an overview of the theoretical frameworks that guided this research. More specifically, the reader will be introduced to a review of motor learning theory and its related constructs, including distributive practice, deliberate practice, and a model of skill acquisition. This will be followed by a discussion on Experiential learning and the related construct of Mentoring and Coaching as it relates to professional development and the literature surrounding video-based coaching in surgical training programs will be explored. Finally, a discussion on self-assessment will be provided in relation to its ability to assist with technical skill acquisition.

Chapter three is devoted to the methodological structure of this research, including the research design, sampling, participants, and instruments employed for data collection. This chapter will also address the research procedures employed, a discussion

on the validity of the instrument used in this research, as well as the materials and

methods used for data collection and analysis.

Chapter four will focus on the presentation of each research question with descriptive and

inferential statistical analysis.

Chapter five will include an in depth discussion and interpretation of the findings, as well

as the conclusion, limitations to the study, and suggestions for future research.

Chapter Two: Review of Related Literature

For this study, in order to properly address the stated problem, help guide the research design, and answer the research questions, I will explore the theory surrounding technical skill acquisition as well as the related constructs that support psychomotor development. This review will better help the reader place these theories in the context of the second-year junior surgical residents who participated in the vascular anastomosis simulation workshop around which this research project was designed. I will first provide a brief overview of motor learning theory, as this forms the basis of understanding human psychomotor development. This will be followed by a discussion on the related constructs of technical skill acquisition, including distributive practice, deliberate practice, as well as a model of skill acquisition and methods to facilitate learner movement through this model. These constructs will all be defined, described, and discussed in relation to surgical education. This will include a discussion on experiential learning, mentoring, and coaching. Finally, I will discuss the current available literature related to video analysis and coaching in the surgical arena and provide an overview of self-assessment and its role in performance improvement.

The limitations of professional ability, and the factors that contribute to an individual becoming an expert in a particular domain, have been widely researched since the 19th century, starting with Sir Francis Galton. In Galton's published work, titled Hereditary Genius (1869), he correctly presented evidence that body size and height was genetically predetermined. In addition to this, however, Galton's theory incorrectly proposed that other innate mechanisms, such as exceptional performance, must be also transmitted from parents to their children. Galton acknowledged that while one can

improve performance through practice, there was a ceiling or a limit to their abilities, regardless of how much practice they put in, because it was dictated by their genetic makeup. Galton emphasized that rapid performance improvements are only evident when one begins training, with subsequent gains becoming increasingly smaller until "maximal performance becomes a rigidly determinate quantity" (p. 15). Galton's theory widely became known as the *Nature* versus *Nurture* debate and, due to lack of research on motor learning in the early 20th century, this belief made up the mainstay of popular culture until the middle of the 20th century (Ericsson, Krampe, & Tesch-Römer, 1993). The literature has shown that there are a host of other factors that contribute to performance excellence, including motivation and repeated exposure (Ericsson et al., 1993), coaching (Ericsson, 2004), and situational practice (Chase & Simon, 1973). Before I address these concepts, particularly as they relate to this research, a discussion on basic motor learning theory is required, as it forms the foundation of any research into human psychomotor development.

Motor Learning Theory

Technical skills can be defined as "the ability to use surgical instruments in an effective and efficient manner and includes economy of motion and safe tissue handling" (Mellinger et al., 2017 p. 52). To help residents attain this level of motor competency, surgical educators should have an understanding as to how humans process information in order to produce long-term changes in their skilled behavior. Theories pertaining to motor learning have been in development since the late 19th century, starting with investigations into sending and receiving Morse code (W. L. Bryan & Harter, 1897, 1899), and the acquisition of typing (Book, 1925). This research led to the explosion of

behavioral psychologists including Thorndike's investigation into knowledge of results (Thorndike, 1927), Hull's classic work on conditioning (Hull, 1943), and Skinner's research on reinforcement (Skinner, 1965). The research presented in the following chapters, however, is grounded in two theories that were put forth by Jack Adams (Adams, 1971) and Richard Schmidt (Schmidt, 1975), and developed in the later part of the 20th century.

Jack Adams' closed loop theory of motor learning (1971, Appendix A) was built on the premise that a simplified stimulus-response behavioral theory, which had been developed using animal models only, was insufficient to explain skilled motor behavior acquisition in humans. Adams took particular issue with the fact that in order for a motor response to occur, some outside stimulus was required. This open-loop model, then, meant that as long as the motivational and habitual states remained constant, the same response (movement) would always be produced (p. 117). Adams realized that this model did not account for the effect feedback can have on the completion of a motor task, particularly feedback regarding error. As such, Adams' closed loop theory of motor learning was the first to be built around the premise of an empirical reinforcing event, or "knowledge of results" (Adams, 1976, p. 88). As defined by Adams, knowledge of results is "information, and how the subject will transform and use it will depend on the type and accuracy of knowledge of results and on the kind of motor task" (p. 90). In contrast to an open loop motor system, which, according to Adams, "has no feedback or mechanisms for error regulation" (p. 89), a closed loop system includes "feedback, error detection, and error correction as key elements" (p.89).

In addition to this construct, Adams describes two states of memory which have a determining role in any motor task, labeled as the memory trace and the perceptual trace. The memory trace is responsible for initiating a movement, choosing initial direction and determining the early portions of movement for the criterion motor objective (Adams, 1971, p.125). The perceptual trace, on the other hand, is the reference mechanism and resulting movement that is developed from past experience and used to guide movement to the correct location along the prescribed pathway (p. 124). It is this perceptual trace, according to Adams, which allows an individual to become more accurate and confident, over time, in carrying out a motor response.

Acknowledging some limitations of Adams' theory, such as needing to first experience a correct location in order to accurately move to that location, one's ability to continue to learn without knowledge of results, and the lack of adaptability to open-loop systems (changing environment), Richard Schmidt (1975) proposed his Schema Theory on motor learning (Appendix B). Borrowing a concept that was originally introduced by Head (1925) in the Psychology discipline, a schema can be defined as "a rule, developed by practice and experience across a lifetime, which described a relationship between the outcomes achieved on past attempts at running a program and the parameters chosen on those attempts" (Schmidt, 2003, p. 367). More specifically, schema theory is based on the formation of a generalized motor program (GMP) for motor learning. Schmidt, Lee, Winstein, Wulf, & Zelaznik (2018) define a GMP as "a motor program for a particular class of actions that is stored in memory and that a unique pattern of activity will result whenever the program is executed" (p. 199). These patterns of activity are then further broken down into a recall schema, which is concerned with the production of movement,

and a recognition schema, in which a "person can estimate the sensory consequences that will occur if that movement outcome is produced," (p. 383), and is analogous to the perceptual trace as defined by Adams.

Perhaps most notable about Schmidt's theory is his description of the various types of feedback that exist, and the role knowledge of results (KR) and knowledge of performance (KP) plays on motor learning. According to Schmidt (Schmidt & Wrisberg, 2004), feedback itself can be divided into two discreet forms, intrinsic and extrinsic. While intrinsic feedback arises from the sensory information gathered from performing a task, it can either be proprioceptive (sensory information received from within the learner's body such as joints/muscles/tendons) or exteroceptive (sensory information acquired primarily through vision and hearing) in nature (p. 277). Extrinsic feedback (also referred to as augmented feedback), however, refers to information received from an outside source, in addition to the intrinsic feedback the learner has already received (p. 279). It is this type of feedback, according to Schmidt, that can affect one's performance, either through KR or KP.

Schmidt put forth a definition of KR which is slightly different, and more specific, than what Adams (1971) proposed. According to Schmidt, KR is "verbal (or verbalizable), terminal (i.e. post-movement) feedback about the outcome of the movement in terms of the environmental goal" (Schmidt et al., 2018, p. 343). Important to note in this definition is that KR, according to Schmidt, does not describe the performance outcome, but simply tells the learner whether he or she has achieved the goal of the performance. Schmidt makes two points about this type of KR. First, it is usually the only meaningful source of outcome information available to the learner; and

second, KR is typically administered in verbal form as terminal feedback (opposed to concurrent) (Salmoni, Schmidt, & Walter, 1984).

This form of feedback is different from KP, which Schmidt defines as "augmented information about the movement pattern the learner has just made" (Schmidt et al., 2018, p. 438). It is this type of feedback which is typically given by instructors or coaches to correct improper movement in learners, and can be given both verbally, as well as through nonverbal means, such as video replay. The fundamental concepts of Adams' closed loop model of motor learning and Schmidt's Schema Theory are paramount to the research that follows and provide a framework for understanding the following concepts which ultimately contribute to technical skill acquisition in junior surgical residents.

Distributive Practice

The concept of spaced versus massed training has been looked at considerably over the last century and a half (See Ebbinghaus, 1885/1964; Hull, 1943 for historical reference). Massed training conditions are defined as "those in which individuals practice a task continuously without rest," whereas spaced training conditions are defined as "those in which individuals are given rest intervals within the practice session" (Donovan & Radosevich, 1999, p. 795). The majority of this, and current, literature supports the notion that short practice sessions, with appreciable time intervals between sessions (distributive practice), leads to better acquisition and retention of skill than when a task is practiced in a continuous (massed) block (Donovan & Radosevich, 1999; Lee & Genovese, 1988). While interest and potential applications of this research remained strong in the educational literature (Adams, 1987), only over the last fifteen years has the

surgical world begun to embrace this concept, and evaluate the outcomes of distributive

training with regard to surgical skill acquisition. Moulton et al. (2006) looked at massed

versus distributive training with junior surgical residents attending a microvascular

anastomosis lab and found a significant difference in both skill retention and competency

for those residents attending the distributive training. Gallagher et al. (2012) also adapted

this style of distributive training to a virtual reality simulation curriculum for

laparoscopic cholecystectomy skill acquisition. The authors were able to show better

rates of performance improvement, and a statistically significant decrease in errors made,

among the distributive training group. Even more importantly, however, was the author's

finding that once a novice acquired skills through laparoscopic training, significant

degradation of that skill occurred after 2 weeks of nonuse.

These studies within surgical training reinforce the concept that retention of motor

skill appears to be task dependent and is influenced by the training interval. There are a

few reasons why distributive practice is preferable when compared to a massed training

schedule, especially within the realm of surgical skill development. First, incorporating

breaks in practice allows for mental rehearsal in anticipation of the next scheduled skills

session, which has been shown to foster the formation of neurologic changes in the brain

that accompany motor skill development (Hall, 2002). Second, there is considerable

mental and physical fatigue associated with skills training. This is especially true during

the early phase of skill acquisition, when the high mental demands of learning the task

can induce fatigue and can interfere with the cognitive aspect of learning a new task

(Tsuda, Scott, Doyle, & Jones, 2009). Fatigue from sleep deprivation has also been

shown to significantly impair psychomotor and cognitive skill development in surgical

residents (Kahol et al., 2008). Perhaps even more significant is the fact that Branscheidt et al. (2019) has shown that the negative effects fatigue has on motor skill acquisition can even extend to subsequent practice days in the absence of fatigue.

Lastly, the benefits of longer intersession training intervals are consistent with the hypothesis of memory consolidation theory. Originally proposed by Müller and Pilzecker (1900) over a century ago, this theory suggest that learning does not create instantaneous, permanent memories. Rather, our memories remain in a state that is vulnerable to disruption immediately after learning a new task, and take time to become fixed (or consolidated) (Lechner et al., 1999). Periods of inactivity, rest or sleep, after a practice session, have also been found to play a significant role in the consolidation of long-term memory (Shea, Lai, Black, & Park, 2000). Perhaps even more importantly, is that this process is not only time dependent (Brashers-Krug, Shadmehr, & Bizzi, 1996; McGaugh, 2000) but also sleep dependent, where a delayed stage of learning occurs without practice, as a result of post-training consolidation. In essence, the literature supports the notion that longer sleep durations yield greater improvements in retention, particularly for procedural memories (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Stickgold, James, & Hobson, 2000; Stickgold, Whidbee, Schirmer, Patel, & Hobson, 2000; Walker et al., 2003). Karni et al. (1994) and Stickgold, Whidbee, et al. (2000) further showed that this process was unable to take place during an interruption in rapid eye movement (REM) sleep, which ultimately prevents overnight performance gain. While this concept of distributive practice is an important component of skill acquisition, there are other factors involved which are also necessary to improve one's skill.

Deliberate Practice

        Caffarella defines transfer of learning as "the effective application by program participants of what they learned as a result of attending an education or training program" (Merriam, Caffarella, & Baumgartner, 2007, p. 211). For surgical residency training programs, successful transfer of learning is a necessary requirement if we are to provide surgical residents with the knowledge needed to perform a technical skill. While distributive practice can ultimately improve these learned skills through spaced trials and memory consolidation, practice and repetition are not the only requirements for achieving expertise in a particular skill. This has been shown in surgical literature, where practice and repetition alone was not enough for novices to ever reach competency in arthroscopy and laparoscopy, despite repeated, sustained practice attempts (Alvand, Auplish, Gill, & Rees, 2011; Grantcharov & Funch-Jensen, 2009). These findings suggest, and are consistent with, newer research on skill acquisition, that there is another element responsible for performance improvement.

        Contrary to Galton's theory of inherent physical talent, Psychologist K. Anders Ericsson suggests that experts are not born, but rather made through a process called deliberate practice. In conducting research on athletes, chess players, and musicians, elite performers are almost always introduced to their skill at an early age, reach peak career performance in their mid to late 20's, and need at least ten years of intense performance before they can reach expert, international level (Bloom & Sosniak, 1985; Chase & Simon, 1973; Ericsson, 2004). Expert performance in a domain is a result of practice and one's deliberate engagement and choosing activities that improve and maintain high performance (Ericsson et al., 1993). While this research has become popularized in

today's culture as the 10,000 hour practice rule (Gladwell, 2008), this does not fully explain deliberate practice. Ericsson's original research (1993) finding was that consistent, gradual improvement in performance was only met under the following conditions; participants were instructed to improve some aspect of their performance on a well-defined task; participants were motivated to improve; participants received detailed immediate feedback on their performance; and participants were provided ample opportunities to improve their performance gradually through repeated practice. Only when these conditions are met, can the term "deliberate" be used to define practice.

With regards to surgery, specifically, while regular practice of a skill has been shown to be an important determinant of patient outcome (Halm, Lee, & Chassin, 2002), it does not fully account for skill level in surgeons. Using simulation training, research by Palter & Grantcharov (2014) as well as Hashimoto et al. (2015) have shown that current surgical training could be improved and that residents can and do reach a higher level of expertise in a skill through the implementation of deliberate practice. This concept has significant implications for surgical residency training, where, each day, novice residents are taught new technical skills that require concentration and repetition (distributive practice), in an effort to reach an acceptable level of everyday performance. According to Ericsson, Nandagopal, & Roring (2009), this process typically takes about 50 hours, after which performance becomes automated, and individuals no longer seek to modify or improve their behavior, which leads to a stable performance plateau. This process emphasizes the fact that if one wishes to attain mastery or expert level of performance in a particular domain, deliberate practice is critical to this process. Before discussing this further, however, it is important to define the levels of skill acquisition

that a novice learner may progress through, on their way to achieving expert level of performance.

Dreyfus and Dreyfus Model of Skill Acquisition

Based on their study of chess players, air force pilots, and army tank drivers and commanders, brothers Stuart and Hubert Dreyfus developed their model of skill acquisition (Dreyfus & Dreyfus, 1979, 1980). This developmental model, based on situated performance and experiential learning, posits that students pass through five levels of skill acquisition through formal education and practice. These levels include: novice, advanced beginner, competent, proficient, and expert (Dreyfus, 2004, Appendix C). According to the Dreyfus brothers, this theoretical model represents a

progression from analytic behavior of a detached subject, consciously decomposing his environment into recognizable elements, and following abstract rules, to involved skilled behavior based on an accumulation of concrete experiences and the unconscious recognition of new situations as similar to whole remembered ones (Dreyfus & Dreyfus, 1986), p. 35).

Since its initial publication, the Dreyfus model has been adapted for use in virtually all disciplines, but has had substantial adaptation within the disciplines of education (Berliner, 2004), dentistry (Lyon, 2015), nursing (Benner, 2001), and medicine (Carraccio, Benson, Nixon, & Derstine, 2008). This model has particular significance to technical skill acquisition in the surgical discipline as, historically, competency in technical skills has simply been assumed after protracted exposure to procedures over the course of five years of training. The problem with this method, however, is that procedural experience in current surgical training is varied, unsystematic and

unstructured (Lodge & Grantcharov, 2011). Furthermore, according to Szasz et al.

(2015), there is a lack of agreement in what defines "competency," and this term is

frequently used interchangeably with "proficiency" within the surgical discipline.

Merriam Webster defines competence as "the quality or state of having sufficient

knowledge, judgment, skill, or strength (as for a particular duty or in a particular

respect)," while proficiency is defined as the "advancement in knowledge or skill"

(Merriam-Webster Dictionary, 2018.). Incorporating the following stages of the Dreyfus

model into practice can provide surgical educators with a more critical understanding of

the differences in these terms and the specific stages of skill acquisition that surgical

residents may progress through in residency training.

     Novice. This initial stage of skill acquisition is characterized by the recognition

of discrete facts and features, which are clearly and objectively defined, and remain

virtually context-free (Dreyfus & Dreyfus, 1986). A novice learner treats each situation

as a new one and looks for appropriate rules to follow. For this level of skill, Dreyfus

notes that it is important to deconstruct the task that is to be taught, as "the student needs

not only the facts but also an understanding of the context in which that information

makes sense" (Dreyfus, 2004, p. 177).

     Advanced Beginner. Through significant practical experience, a novice learner

can progress to an advanced beginner stage of skill acquisition. While these learners still

demonstrate marginally acceptable performance, they are able to perceive similarities of

concrete situations and can relate them to previous examples of the same experience

(Benner, 2001). According to Dreyfus and Dreyfus (1986), practical experience is much

more productive than any amount of verbal description at this stage. This learner, like a

novice, can take in little of the situational aspects of performance, as they continue to "concentrate on remembering the rules they have been taught" (Patricia Benner, 1982, 404).

Competent. The competent learner is able to pull from their past situational experience to create a problem-solving format which allows them to adopt a hierarchical process of decision-making Dreyfus and Dreyfus (1986). More specifically, a "competent performer seeks rules and reasoning procedures to decide which plan or perspective to adopt," in order to prevent mistakes (Dreyfus, 2004, p. 178). In this respect, while they may lack the efficiency of a more experienced performer, they are developing the ability to manage situational contingencies (Benner, 2001). It is important to note in this stage that, unlike the novice or advanced beginner, who act according to strict rules, the competent performer becomes more vested in the resultant outcomes based on their actions (Dreyfus & Dreyfus, 1986).

Proficient. A learner at the proficient stage of performance relies less on rules, is more emotionally involved in their tasks, and has the ability to discriminate among a variety of situations in order to choose the most effective action to accomplish their task (Dreyfus, 2004). More specifically, a proficient performer "while intuitively organizing and understanding his task, will find himself thinking analytically about what to do" (Dreyfus & Dreyfus, 1986, p. 29). In essence the situation is what ultimately guides the proficient performer's response (Benner, 2004).

Expert. This final stage of acquisition is marked by full engagement and efficiency on the part of the performer. According to Dreyfus, an expert "not only sees what needs to be achieved…he or she also sees immediately how to achieve this goal"

(2004, p. 179-180). A curious distinction here is that responses at this stage in performance are reactive rather than studied and premeditated. Dreyfus & Dreyfus (1986) note that "when things are proceeding normally, experts don't solve problems and don't make decisions; they do what normally works" (Dreyfus & Dreyfus, 1986, p. 31;

This five-stage model will be instrumental in guiding this particular research project as I employ a variety of principles of adult education to assist in the technical skill acquisition of junior surgical residents participating in a vascular anastomosis simulation lab. While this model has been widely adopted, it is not without its criticisms. Hargreaves and Lane (2001) have criticized the fact that a linear model of skill acquisition is unable to sufficiently explain the everyday experiences of learning. Perhaps even more pertinent, as it pertains to social constructivism, the model has been criticized for its apparent lack of social structure or social knowledge (Purkis, 1994; Rudge, 1992). It is this point that needs further exploration, as the Dreyfus model does not seem to address how to encourage or assist learners through the stages of skill acquisition. The role of reflection is also minimized in this five-stage model and needs to be accounted for as it pertains to this research. Addressing these criticisms and concepts are crucial in assisting technical skill acquisition. Therefore, the experience that ultimately shapes learning needs to be reviewed, and the role of a more experienced other, also referred to as a coach, must be defined in relation to this construct. To help bridge these constructs, a brief discussion of Experiential learning theory (Kolb, 2014) is required.

Experiential Learning

Dewey (1938) was one of the first theorists to recognize the role experience plays on education, stating "all genuine education comes about through experience" (p. 13). Heavily influenced by this and the earlier work of Lewin (1935) and Piaget (1980), David Kolb further expanded on this concept, defining learning as "the process whereby knowledge is created through the transformation of experience" (Kolb & Kolb, 2005, p. 194). Adult educator Malcolm Knowles originally defined Andragogy as "the art and science of helping adults learn" (Knowles, 1980, p.43), and Kolb's Experiential Learning Theory (ELT) is now considered by many to be a fundamental theory in Andragogy. Emphasizing "the central role that experience plays in the learning process" (Kolb, Boyatzis, & Mainemelis, 2001, p. 2), the theory itself is built upon the following six propositions: learning is best conceived as a process, not in terms of outcomes; all learning is relearning; learning requires the resolution of conflicts between dialectically opposed modes of adaptation to the world; learning is a holistic process; learning results from synergetic transactions between the person and the environment; and learning is the process of creating knowledge (Kolb & Kolb, 2005).

Kolb (2014) further conceptualized that in order to learn from an experience, four different types of abilities are required: 1) a willingness to involve oneself in a new experience (concrete experience); 2) observation and reflection in order to view the experience from a variety of perspectives (reflective observation); 3) the ability to analyze one's actions so new ideas and concepts can be created from one's observations (abstract conceptualization); 4) The ability to problem-solve and make decisions, so new ideas and concepts can be used in actual practice (active experimentation).

According to Kolb, these adaptive learning abilities combine to form two dialectically related modes of a grasping experience (concrete experience and abstract conceptualization) and two dialectically related modes of a transforming experience (reflective observation and active experimentation), and it is up to the learner to "continually choose which set of learning abilities he or she will use in a specific learning situation" (Kolb, Boyatzis, & Mainemelis, 2001, p.3). ELT is typically depicted in an idealized learning cycle, and students are strongly encouraged to use all four components in order to enhance their learning (Appendix D). Important in this learning cycle is the assumption that the more often a task is reflected upon, there exists a greater opportunity for the learner to modify and refine their efforts.

It is this research that also led Kolb to propose a Learning Style Inventory (LSI) which identified four different learning styles associated with varying approaches to learning: Diverging learners, whose dominant learning abilities are Concrete Experience (CE) and Reflective Observation (RO); Assimilating learners, whose dominant learning abilities are Abstract Conceptualization (AC) and Reflective Observation (RO); Converging learners, whose dominant learning abilities are Abstract Conceptualization (AC) and Active Experimentation (AE); and Accommodating learners, whose dominant learning abilities are Concrete Experience (CE) and Active Experimentation (AE) (Kolb et al., 2001).

To put these learning styles into further perspective, Divergers are usually open minded, imaginative, and prefer to work in groups; Assimilators prefer reading, lectures and analytical models; Convergers are apt to experiment with practical applications of new ideas; and Accommodators tend to prefer hands-on learning experiences (Kaushik,

2017). As a point of reference for this research, this LSI has been looked at considerably in both the medical and surgical literature, as a way to identify the types of learners that make up these residency programs, in an effort to make training more efficient and effective. Adesunloye, Aladesanmi, Henriques-Forsythe, & Ivonye (2008) found that medical residents and faculty have a predominantly assimilating learning style, while research on surgical residents and faculty have found a predominantly converging type of learning style (Engels & de Gara, 2010; Mammen et al., 2007). Understanding the specific learning style of our trainees will allow educators, or a more experienced other, develop specific curriculums and programs which may better assist a learner achieve their full potential when learning a new skill. This task can best be conceptualized through formalized mentoring and coaching.

Mentoring and Coaching

Much like the criticism the Dreyfus model of Skill acquisition receives with regards to reflection (Peña, 2010), Kolb's ELT (2005), in its original form, has also been criticized for minimizing the role reflection plays in the learning process. Boud et al (1985) notes that while Kolb's model "has been useful in assisting us in planning learning activities and in helping us to check simply that learners can be effectively engaged," they comment, "it does not help… to uncover the elements of reflection itself" (p.13). Philosopher Donald Schön (1983) describes the process of reflecting-on-practice as a way for practitioners to "think back on a project they have undertaken, a situation they have lived through, and they explore the understandings they have brought to the handling of the case" (p. 61). This is an incredibly important aspect of learning a new task, as it allows one to think about their own progress and determine the ways they may

change their practice through experiential learning next time the task is encountered. The question now becomes, is it possible to help transform a learner's experience through this reflective process, and assist them with attaining higher levels of skill over time?

Acknowledging this question, Peno & Silva Mangiante (2012) created the Purposeful Ongoing Mentoring Model (POMM) to help learners move through the stages of skill acquisition through mentoring, scaffolding, and reflection. The authors adapt a definition put forth by Strowd and Reynolds (2013), who characterize a mentor as "a voluntary and active participant in the personal and professional development of the mentee, offering knowledge, experience, guidance, support, and opportunity for advancement" (p. 244). In this model, the mentor's relationship and actions towards the learner are guided by the work of Vygotsky (Vygotsky, 1978) and his concept of scaffolding. Vygotsky described learning as an active process in which learners construct their own understanding and knowledge of the world, through experiencing things and reflecting on those experiences. As part of this description of learning, Vygotsky described a zone of proximal development (ZPD) or "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers" (Vygotsky, 1978, p. 86).

By identifying a learners ZPD and employing the use of scaffolding, the mentor, or more capable peer "provides a model for a higher level of practice through demonstration and/or explanation (challenge) while supporting (coaching/feedback) the learner's attempts to make sense of and emulate what is being taught" (Peno, Silva Mangiante & Kenahan, 2014, p. 6). Finally, the POMM incorporates the work of Donald

Schön (2017) to promote the art of reflection on the mentee in an effort to "illuminate even subtle differences that need alteration" (p. 7), in an effort to improve practice. According to the authors, the POMM model is "meant to supply a frame for purposefully thinking about, preparing for, and developing actions for the acquisition of higher levels of skill in any area of work" (Peno & Silva Mangiante, 2012, p. 13). While the POMM allows us to better understand how the relationship between a learner and a more experienced other can facilitate the reflective process and promote their movement through the Dreyfus model, I must differentiate between a mentor and a coach, as it pertains specifically to this research.

Coaching author Timothy Gallwey defines coaching as "the art of creating an environment, through conversation and a way of being, that facilitates the process by which a person can move toward desired goals in a fulfilling manner" (Gallwey, 2002, p.177). These goals can further be defined as "a cognitive image of an ideal stored in memory for comparison to an actual state; a representation of the future that influences the present; a desire (pleasure and satisfaction are expected from goal success); a source of motivation, an incentive to action" (Cochran & Tesser, as cited in Street, 2002, p.100). According to Grant (2001), the difference between a Mentor and a Coach is usually based on the objective. In coaching, the objective is skill development and performance enhancement, while in mentoring, the objective is typically long-term career development. Grant also notes that while "mentoring traditionally involves an individual with expert knowledge in a specific domain passing on this knowledge to an individual with less expertise," the coach does not necessarily need to be an expert in the trainees

area of learning, needing to "only have expertise in facilitating learning and performance enhancement" (p. 30).

Further distinctions between the differences between coaching and mentors have been adapted from Passmore (2007) and can be found in (Appendix E). For the purposes of this research, as the relationship between the facilitator and the learners took place over the course of five weeks, with a specific focus on technical skill improvement, I will refer to this relationship as "coaching," to stay aligned with the definitions provided. Aside from this change, the goal of this research remains the same. In an effort to improve performance, I will employ the use of coaching, scaffolding and reflection on practice, as described in the POMM, to assist our learners in achieving a higher level of skill acquisition. Having described this concept, I now turn our attention to the current available literature related to video analysis and coaching in the surgical setting.

Video Analysis and Coaching in the Clinical Setting

In the clinical setting, feedback has been defined as "specific information about the comparison between a trainee's performance and a standard, given with the intent to improve the trainee's performance" (Van De Ridder et al., 2008). Soon after video cameras became widely available, video replay for analysis became established as an effective method for performance review, providing feedback, and contributing to self-assessment in both team and individual sports (Lounsbery & Sharpe, 1996; Winfrey & Weeks, 1993). Early adoption of video analysis, combined with some type of verbal feedback, as a method to enhance performance improvement in surgical training programs, however, failed to replicate these perceived benefits. Working with surgical residents in an orthopedic simulation lab, Backstein, Agnidis, Regehr, & Reznick (2004)

were unable to demonstrate any significant improvement in technical skills when video feedback analysis was incorporated into their model. To assess whether this failed improvement in technical skill was due to participants receiving only a single exposure to video feedback analysis, the main authors of this research performed a second study using a vascular anastomosis bench model. Even when surgical residents were exposed to repeated video feedback analysis over the course of three weeks, the authors, again, failed to show any improvement in technical skill (Backstein, Agnidis, Sadhu, & MacRae, 2005).

More recent research, however, has begun to show some promising results when video feedback analysis was incorporated into their training programs. During a laparoscopic suturing workshop, for instance, Jamshidi, LaMasters, Eisenberg, Duh, and Curet (2009) found significant improvement in knot tying skill in residents who participated in their own video analysis of previous attempts. Similarly, when Farquharson, Cresswell, Beard, and Chan (2013) compared verbal feedback alone to verbal feedback plus video analysis in a group suturing workshop, medical students who participated in the video analysis group were significantly better in their suturing technique, specifically with regards to needle handling, hand-eye coordination, and instrument familiarity. Karem et al. (2015) evaluated orthopedic residents' use of head mounted video recording during an articular fracture simulation lab. The authors found that the intervention group, who reviewed their video with a senior orthopedic faculty member prior to their next attempt, significantly performed at a higher level than the control group. This particular research highlights an important concept that may be

responsible for augmenting video analysis, that is, the specific type of feedback and analysis provided.

Nesbitt, Phillips, Searle, and Stansby (2015) evaluated the effect of individualized video feedback with expert analysis, unsupervised video feedback, and standard lecture format in 35 medical students during a suturing workshop. While they found a significant difference using video feedback over standard lecture format, they ultimately concluded that students could attain similar levels of technical skill acquisition using either unsupervised video feedback or individualized video feedback with expert analysis. More recently, Phillips et al. (2017) evaluated the effectiveness of unsupervised video feedback versus direct expert feedback during a skills lab on intravenous cannulation, catheterization, and suturing. While they also found utility in video-assisted feedback, they failed to show a significant difference using direct expert feedback. These particular studies, however, also failed to ensure the feedback provided by faculty remained consistent across individuals and groups. Because of these inconsistencies, the use of surgical coaching, combined with video analysis, may be more valuable for helping surgical trainees improve technical skills, as it provides faculty with a standardized format to analyze and critique individual performance.

In a randomized controlled study evaluating the effect of video-based coaching on laparoscopic surgical skills, Singh, Aggarwal, Tahir, Pucher, & Darzi (2015) found that students exposed to video-based coaching experienced enhanced laparoscopic surgical performance compared to those who only participated in online tutorials and practice sessions. In a similar study, comprehensive surgical coaching, including video feedback analysis, was evaluated in surgical residents performing minimally invasive gastric

bypass surgery Bonrath et al. (2015). Over the course of their eight-week rotation, the

authors concluded that residents exposed to comprehensive, video-based coaching

demonstrated improved technical skill, significant error reduction, and improved self-

assessment compared with conventional training. More recently, a randomized

controlled trial assessing surgical resident's ability to perform a side-to-side intestinal

anastomosis on cadaveric dog bowel was also able to show a significant improvement in

technical skills in the experimental group exposed to video-based coaching (Soucisse et

al., 2017). Based on these studies, it appears that when structured coaching was

instituted, in addition to video playback analysis, improvements in technical skills were

observed when compared to control groups.

Self-Assessment

Self-Assessment has been defined as "a process of personal reflection based on an

unguided review of practice and experience for the purposes of making judgments

regarding one's own current level of knowledge, skills, and understanding as a prequel to

self-directed learning activities that will improve overall performance and thereby

maintain competence" (Eva & Regehr, 2007, p. s81). The importance of self-assessment

has been vigorously researched over the last thirty years in medicine, and has been

identified as a crucial aspect of professional self-regulation (Arnold, Willoughby, &

Calkins, 1985; Boud, 1999; Gordon, 1991). As noted by Eva & Regehr (2005),

however, most of these studies have casted doubt on an individual's ability to adequately

perform self-assessment, with a lot of the conclusions being that overall, it is quite poor.

One potential explanation for this has pointed to the methodological weaknesses of these

studies and their ability to adequately evaluate self-assessment (Ward, Gruppen, &

Regehr, 2002). Research does show, however, that the potential of self-assessment is twofold; It can function as an identifier of one's weaknesses in a task and can also help identify one's strengths (Ross, 2006). Bandura's social cognition theory also recognizes that students who perceive themselves to be successful on a task are more likely to believe they will be successful in the future (Bandura, 1997). This is a powerful concept with regards to learning a new task.

It has long been shown, however, that learners tend to be poor evaluators of their own performance. Kruger & Dunning (1999) performed research on psychology students and found that low performers are unconsciously incompetent and tend to overestimate their ability, while high performers underestimate their ability. These results were also replicated in medical students (Bryan, Krych, Carmichael, Viggiano, & Pawlina, 2005) and medical residents (Parker, Alford, & Passmore, 2004). When Hu, Tiemann, & Brunt (2013) evaluated medical students and surgical interns performing suturing and knot tying tasks, they found that novice trainees over-estimated their basic technical skills when compared to their assessment by a senior surgeon. With regards to surgical residents specifically, Herrera-Almario et al. (2016) evaluated the self-assessment scores of third year general surgery residents using video playback analysis of their laparoscopic skills. The authors found that residents consistently scored themselves lower than faculty scoring both before and after video analysis of their performance.

What is important to point out, especially in terms of how it relates to this research, is that methods to improve self-assessment in learners has been looked at quite extensively. In a systematic review by Colthart et al. (2008), the authors found that self-assessment can be enhanced by feedback, particularly video feedback and verbal

feedback, and by providing specific criteria with regards to assessment and benchmarking guidance. This finding of video feedback improving self-assessment is not a new concept, and, in fact was studied specifically by Martin, Regehr, Hodges, & McNaughton (1998). Using videotaped benchmarks of interviewing skills with family medicine residents, the authors were able to show a significant improvement in the resident's self-assessment ability after viewing the video of the benchmark and then their own performance. This was also replicated with medical students and a suturing task in a simulation environment. When Hawkins, Osborne, Schofield, Pournaras, & Chester (2012) evaluated thirty one medical students on this task, they found no difference in self-assessment scores after watching their own video playback. This was significantly different from the self-assessment scores which occurred after video feedback combined with a video benchmark performance. It appears that, when compared against a benchmark, video analysis allows individuals the ability to identify their own strengths and weaknesses in the context of good professional practice.

Chapter Two Summary

This chapter provided an in-depth discussion of the theoretical frameworks and concepts that guided this research project. A brief historical overview on the limitations of human professional ability and the factors that contribute to performance excellence was discussed. In order to provide a basis for understanding human psychomotor development, an overview of two major theories of motor learning, Jack Adams closed loop model of motor learning, and Richard Schmidt's Schema Theory on motor learning was discussed. Both of these theories attempt to explain skilled motor behavior

acquisition in humans, in an attempt to move beyond the simple stimulus-response theories that had predominated during the early twentieth century. While Jack Adam's theory introduced the concept of refining a motor act through "knowledge of results," Richard Schmidt's theory further described the difference between acting on "knowledge of results" and "knowledge of performance," in relation to what he termed the generalized motor program (GMP).

Moving on from motor theory, the chapter described the concepts of distributive and deliberate practice as they relate, not only to learning a new skill, but to becoming an expert performer in a particular domain. This concept introduced the 10,000-hour practice effect, but further delineated that in order to achieve gradual, consistent improvement in performance, a combination of motivation, feedback, and repeated practice were needed. In order to identify mastery in performance, however, it was suggested that the Dreyfus model of skill acquisition be introduced to help guide instruction. This developmental model is based on situated performance and identifies five levels of skill; Novice, Advanced Beginner, Competent, Proficient, and Expert.

In order to explore the process of learning through one's experience, the major tenants of Kolb's Experiential Learning Theory (ELT) were explored, including an overview of Kolb's learning style inventory (LSI) which can be helpful in identifying the particular learning style of a student. Noting some of the criticisms of ELT, particularly its lack of emphasis on reflection, the definitions of coaching and mentoring were then explored as a method to help support reflecting on learning. Tying these concepts together, the Purposeful Ongoing Mentoring Model (POMM) was introduced as a way to assist learners improve performance with the help of a more knowledgeable other and

guide them through the stages of skill acquisition through mentoring, scaffolding, and reflection. A review of the literature on use of video analysis and structured coaching methods was then discussed in relation to their current usage in surgical residency training programs. The chapter concluded with an overview of the definition of self-assessment and the methodologies that can enhance a learner's ability to accurately assess their performance in an effort to improve practice. The attention will now turn towards a detailed description of the methodology that formed the basis of this research.

Chapter Three: Research Design and Methodology

This study was developed to compare self-assessment video analysis against structured, individualized, video-based coaching as a way to improve technical skill acquisition in junior surgical residents participating in a vascular anastomosis simulation workshop. Based on this concept, the following research questions were used to help guide this research:

RQ1A: Based on expert evaluator scoring using a validated technical skills checklist, to what extent do residents who participate in structured, individualized, video-based coaching attain technical skill, compared to residents in a self-assessment, video analysis only group?

RQ1B: Using the validated technical skills checklist, to what extent do resident's self-assessment scoring of their own technical skill differ between those who participate in the structured, individualized, video-based coaching group, compared to those in the self-assessment, video analysis only group?

RQ2: Based on expert evaluator scoring using the validated technical skills checklist, do residents who participate in structured, individualized, video-based coaching attain technical skill at a different rate than residents in the self-assessment video analysis only group?

RQ3: Based on the validated technical skills checklist, to what extent do resident self-assessment scores of their technical skill correlate with expert evaluator scores?

RQ4: To what extent do the current posttest scores on time to fashion anastomosis and leak rate compare with the prior three years of resident vascular anastomosis scores?

Research Design

In order to adequately address the research questions, this study employed a quantitative approach using a between groups experimental research design. Quantitative research is defined by Creswell as an "approach for testing objective theories by examining the relationship among variables. These variables, in turn, can be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures" (2014, p. 4). Quantitative designs are most closely aligned with a postpositivist worldview, and adhere to the scientific method which stipulates that "a researcher begins with a theory, collects data that either supports or refutes the theory, and then makes necessary revisions and conducts additional tests" (Creswell, 2014, p.7). A quantitative methodology was chosen for this research due to the specific nature of the research questions. According to Creswell (2014), "if the problem calls for (a) the identification of factors that influence an outcome, (b) the utility of an intervention, or (c) understanding the best predictors of outcomes, then a quantitative approach is best" (p. 20). This randomized controlled research project will be implemented using a vascular anastomosis bench model simulation workshop in a large, academic, Level 1 trauma center in the Northeastern United States. Bench models and simulation training have continued to gain traction over the last decade as a means of improving residents' technical skills. The current literature has shown a positive correlation between bench model training and technical competency in the operating room (Wilasrusmee, Lertsithichai, et al., 2007), and have also been associated with improved patient outcomes (Brydges et al., 2015).

Participants

As part of surgical training, the ACGME requires residents to have access to, and participate in, simulation labs throughout training to "address acquisition and maintenance of skills with a competency-based method of evaluation" (Accreditation Council for Graduate Medical Education, 2019, p. 7). In an effort to help our surgical residents acquire the technical skills necessary to perform a vascular anastomosis, our local residency curriculum requires residents to attend our vascular anastomosis simulation workshop at the beginning of their second year of training, which runs annually from July through August. Because these residents were all required to participate, this study employed convenience sampling by soliciting the entire class of second-year surgical residents scheduled to attend this vascular anastomosis lab (total N=13) for study participation starting in July 2019. As this is a convenience sample, and not a random sample, the resident participants cannot be considered representative of any population (Fraenkel, Wallen, & Hyun, 2011). It is important to note, however, that this research project was conducted using a randomized controlled design. According to Fraenkel, Wallen, and Hyun (2011), "the essential ingredient of a true experimental design is the subjects are randomly assigned to treatment groups" (p.270). All residents enrolled in this research were randomly assigned, using a random number generator, to either the structured, individualized, video-based coaching group (Experimental) or the self-assessment video analysis group (Control). To try to minimize any differences between resident participants from the three departments represented (general surgery, urology, and plastic surgery), the number generator was used three times to assign each department's residents into either the control or experimental group. The facilitator

responsible for running the anastomosis lab and providing the coaching sessions, along with the two expert evaluators, are all local, board-certified faculty from the division of vascular surgery, and all have been practicing for a minimum of five years. See the table below for participant attributes:

Variables

An independent variable is defined as a variable "that (probably) cause, influence, or affect outcomes" (Creswell, 2014, p. 52), while dependent variables are "those that depend on the independent variables; they are the outcomes or results of the influence of the independent variables" (p. 52). With regards to the research design of this study, the independent variable being evaluated has two levels, and includes self-assessment video analysis, and structured, individualized, video-based coaching. The dependent variable that is being evaluated is technical skill acquisition and will be measured by performance scores, using the validated MOSAT and GRS evaluation tool, in addition to total time and leak rate, that were obtained at four separate recorded trials that occurred over seven weeks during the 2019 vascular anastomosis workshop.

Instruments

Technical performance on the vascular anastomosis was assessed using the Mini Objective Structured Assessment of Technical Skills (MOSAT) checklist (Appendix F) combined with the Global Rating Scale (GRS, Appendix G). The MOSAT is a detailed checklist consisting of 24 operation-specific actions necessary to perform the vascular anastomosis that participants were required to produce for this simulation lab. These 24 actions are all graded on a binary scale, where the performer gets either zero versus one point depending on if they successfully performed each action. For purposes of this

54

research, the total possible MOSAT score was cut down from 24 to 22 by eliminating the two variables related to "Control of vessel." This was due to the variation of materials provided to participants for this particular anastomosis lab, specifically the Ethicon suturing jig, which already has built in clamps to secure the Graft. The second part of the scoring evaluation, the GRS, consists of eight specific dimensions, each related to some aspect of operative performance, that is also pertinent to this simulation lab exercise. Each dimension of performance is graded on a 5-point Likert scale with "1" being the lowest possible score and "5" being the highest possible score per category. The highest possible score a participant may receive on the GRS, based on this scoring, is 40.

As this research design is incorporating an already existing instrument, it is important to document the instrument background, as well as its validity and reliability. As defined by Fraenkel et al. (2011), validity "refers to the degree to which evidence supports any inferences a researcher makes based on the data he or she collects using a particular instrument" (p. 148). These authors also define reliability as "the consistency of the scores obtained-how consistent they are for each individual from one administration of an instrument to another and from one set of items to another" (p. 154). The MOSAT checklist and GRS have previously been reported to have an inter-rater reliability coefficient of 0.781 and 0.843 (Reznick, Regehr, MacRae, Martin, & McCulloch, 1997). The combined use of rating scales, such as the MOSAT and GRS, have been implemented in most surgical disciplines over the last two decades, and have recently been deemed the best overall approach to assess technical competence in surgical trainees (Szasz et al., 2015). Based on the validity and reliability of these instruments, their applicability to the required task, and their ability to specifically

address my research questions, the MOSAT and GRS were ultimately selected for use in this research.

In addition to this scoring, all residents were evaluated on their time to perform each vascular anastomosis and all anastomoses were tested for a leak based on a low fidelity model. Total time to perform the anastomosis was calculated based on the time the first stitch was performed, up until the final knot had been tied and cut by the resident. As all of the trials were video recorded, this total time was calculated and documented based off the video session for accuracy. All anastomoses were also tested for a leak rate after each recording session by clamping the distal ends of the graft to isolate the anastomosis, then perfused with normal saline solution for a total time of 30 seconds. This low-fidelity model is similar to the one our institution reported in 2013 (Okhah, Morrissey, Harrington, Cioffi, & Charpentier, 2013). The saline bag was suspended on a pole at a height of 72 inches from the ground. An intravenous line (10 drops/ml, 104in, Baxter) was used to connect the saline to the clamped graft. This clamped graft was then set in a 1000ml graduate cylinder, resting on a table approximately 30 inches off the ground. A stopwatch was then used to time the 30 seconds allotted for saline perfusion. Any leaked saline captured in the 1000ml beaker during the 30 second perfusion was then transferred to a smaller 100ml graduated cylinder. This provided an accurate assessment of total leak (milliliters), which was then documented on the data collection sheet according to the unique resident identification number.

In order to better quantify a resident's technical performance and attempt to identify participants who have achieved baseline competency in an end to side

anastomosis, a scoring system was created by the principal investigator (PI), based upon the variables of MOSAT, GRS, time and leak rate. These variables were combined into a single formula which attempted to place the appropriate weight on the more significant variables (MOSAT and GRS) and less weight on time and leak rate. This score was developed based on the functional skill of novices entering this anastomosis lab and the categorical weight applied to the variables attempts to best represent this. Both the MOSAT and GRS were given a standard multiplier of 2 due to their validity and applicability to the task. The MOSAT, being a technical checklist of the specific actions necessary to perform the anastomosis, was given an additional multiplier of 2. Time was then taken off for total time to complete the anastomosis as well as leak rate. The complete formula that was created for this score can thus be expressed by the following equation:

Combined Technical Score = 2 x (MOSAT x 2) + (GRS x 2) – Time – Leak Rate

Procedures

Residents were randomly assigned into one of two groups, a self-assessment video analysis only group (control) and a structured, individualized, video-based coaching group (experimental). Upon presenting to the lab on week one, all residents were instructed on the goals and objectives of the vascular anastomosis lab. In addition, residents were made aware of the intent to use this lab as part of a research project to attempt to determine the effects of Coaching and Self-Assessment on performance improvement and skill acquisition. All residents were then offered a copy of the informational waiver of consent to read and maintain for their records (Appendix H). Participants were then asked to perform the following task: perform an end-to-side

anastomosis of an 8-mm polytetrafluoroethylene (PTFE) graft (LeMaitre Vascular, Inc. Burlington, MA). This anastomosis was to be performed using a continuous running suturing technique, with a 6-0 double armed polypropylene suture with a C-1 needle (Ethicon). Each participant then moved to their individual stations, with a partner, where a video recording was made of them performing this anastomosis on the PTFE graft. Residents always worked in groups of two, so that as one resident performed the anastomosis, the other resident, sitting across from them, could serve as their assistant.

Each station contained an Ethicon mounting jig, sterile gloves (for deidentification), a 6-0 double armed polypropylene suture with a C-1 needle, and two segments of an 8mm PTFE graft (8cm and 6cm in length). A variable array of instruments was provided, and residents had the ability to choose which instruments they thought were needed to correctly perform this anastomosis (Figure 1). Important to note, the grafts were cut to specified length by the PI, but no graftotomy was created, and grafts were not spatulated prior to the procedure. Graftotomy and spatulation was left up to the discretion of the resident, and this portion of the procedure was not included in the total time for anastomosis.

Figure 1. Photos from the simulation lab of the instruments provided and suturing jig setup.

Immediately following the pre-test in week one, all participants received a link to a private, online, video streaming platform of the vascular faculty member discussing the various types of instruments required to perform the vascular anastomosis, the proper suturing technique that is required, and some helpful hints for anastomosis success. Two videos were ultimately made to address these concepts and they can both be viewed here: **Video 1**, **Video 2**. There were no restrictions on the amount of times these videos could be reviewed by participants.

Residents returned each of the following six weeks to practice and perform this same anastomosis. Each resident was video recorded a total of four times over the course

of the seven weeks; week one (pre-test), then again in week three, week five, with a final (post-test) recording occurring in week seven. Weeks two, four, and six were standard practice sessions and residents were not video recorded at these sessions. All residents received generalized verbal feedback during their anastomosis sessions as per the current standard vascular anastomosis simulation lab protocol (one vascular faculty member for all residents). There were no restrictions on the types of questions regarding performance residents could ask the vascular attending facilitator during these sessions.

Residents in the structured, individualized, video-based coaching group received, in addition to this conventional simulation lab, video-based coaching sessions by the vascular attending, structured around a specific coaching model. These coaching sessions occurred during the non-video weeks (two, four, and six), prior to the practice sessions, for a total of three coaching sessions per resident in the experimental group. For the purposes of this research design, the GROW model of coaching (Whitmore, 2010) was selected due to its simple design structure and targeted focus. Used by more than 40% of coaching psychologists, the GROW model breaks coaching sessions down into four phases; Goals, Reality, Options, and Wrap-up (Grant, 2011, Appendix I). As a tool used for goal-setting and problem-solving, this model employs an inside-out teaching strategy that assumes learners have unlimited potential which can be unlocked with proper coaching. Despite its widespread use, however, it has been criticized for failing to consider where people are coming from prior to helping them get where they want to go (Bishop, 2015). For this reason, Grant suggests adding "RE" to the acronym, which allows for time to review and evaluate, with the premise that "each coaching session should start with a process of reviewing and evaluating the learnings and actions

completed since the last session" (p. 124, Appendix J). This critical aspect of the coaching model allowed the surgical coach to help guide resident reflection through the experiential learning process as described by philosopher Donald Schön, and conceptualized in the POMM.

Video recordings were all deidentified and maintained on a private, online, video streaming platform. Within 48 hours of their recording session, residents in both the control and experimental group were sent a unique email link to their video. This link allowed them unlimited viewing access to assess their technique and perform comparison analysis with the expert video they were provided at the beginning of the anastomosis lab. Upon completion of each recorded anastomosis session (week one, three, and five), residents in both groups were asked to evaluate themselves using the MOSAT and GRS electronic evaluation form.

In addition to the residents performing a self-evaluation of their videos, each video was also sent to the two blinded vascular faculty members for analysis and scoring using the same electronic evaluation form incorporating the MOSAT and GRS. Group randomization in this research design facilitated the faculty blinding technique to better assess if the intervention actually affected the outcome. More specifically, *The Cochrane Handbook* defines blinding as:

> the process by which study participants, health providers, and investigators, including people assessing outcomes, are kept unaware of intervention allocations after inclusion of participants into the study. Blinding may reduce the risk that knowledge of which intervention was received, rather than the intervention itself,

affects outcomes and assessments of outcomes. (Higgins & Green, 2011, Box

8.11.a)

The deidentified videos from each resident trial in weeks one, three, five, and seven were

sent to the blinded faculty evaluators through a unique email link the same week they

were recorded. There was no specific order in which the videos were sent to the blinded

faculty, and the order in which they received each video varied across the sessions.

Validity

Validity can generally be referred to as the level of accountability and legitimacy

that is strived through data collection, analysis and interpretation (Onwuegbuzie &

Teddlie, 2003). As such, it is an important concept in quantitative research because it

refers to the strength of the conclusions that are drawn from the results. More

specifically, are the reported results of the research accurate? There are three types of

validity that researchers need to account for when performing research, specifically

content validity, construct validity, and criterion validity. Content validity is concerned

with the extent a research instrument accurately measures all aspects of the content.

Construct validity refers to the extent an instrument measures the intended construct, and

criterion validity is the relationship between scores obtained using the instrument or

instruments (Fraenkel, Wallen, & Hyun, 2011).

To help establish and maintain content and construct validity, specifically, the

three faculty members from the division of vascular surgery met with PI of this research

in person, one month prior to the start of the anastomosis lab. During this meeting, the

concepts and core objectives of the anastomosis lab were discussed, and the scoring

rubric for the MOSAT and GRS, as a measure to quantify technical skill acquisition, was

agreed upon. All sections from the MOSAT and GRS were discussed at this meeting and all agreed that the components were necessary actions for correctly performing a vascular anastomosis. After this meeting, the vascular surgeon who was to serve as the coach and facilitator throughout the simulation lab, met separately with the PI for a training session on how to implement the coaching model with the experimental group attending the simulation lab. This faculty member was guided on the specific structure and components of the Re-GROW coaching framework and demonstrated understanding of how to best implement this model into practice (Appendix K), and how to best incorporate specific feedback into each resident coaching session.

Data Collection

The data collection for this research came from several sources. The primary source of data collection was through video capture. A mini action camera was able to be mounted to an articulating arm placing the camera's field of view approximately 12 inches above the Ethicon suturing jig the residents used to perform the anastomosis on. The recorded video segments from each anastomosis session were all edited and deidentified by the PI, using video editing software, and then uploaded to a private, online, video streaming platform. From this platform, a unique link was created for each video, which was then securely sent through email to each respective resident and the two blinded faculty for evaluation purposes. Only those with the unique link were able to view the video, and no links to the videos were made public. An example anastomosis video can be viewed here: Example Video**.**

For standardization, security, and ease of data collection, the MOSAT and GRS were combined into one single evaluation form (Appendix L) using our local Research

Electronic Data Capture (REDCap). The REDCap evaluation allowed both residents and

blinded faculty to quickly evaluate the anastomosis video using either a smartphone,

tablet, or personal computer. The evaluations were all maintained on the local REDCap

server, which only the PI had access to. This format allowed the PI to maintain the data

on a local, secure, web-based server throughout the study, and easily export the

deidentified evaluation data into the Statistical Package for the Social Sciences (SPSS)

for data analysis. The MOSAT and GRS evaluation form used for this project's data

collection can be found here: MOSAT and GRS Eval.

Data Analysis

Appropriate inferential and descriptive and inferential statistical methods were

applied based on specific probability distributions for each variable. Results are reported

as frequencies, distributions, means, medians, modes, and standard deviations. For

continuous data (RQ1A), the two groups (self-assessment video analysis only versus

structured, individualized, video-based coaching group) was compared using an

independent T test and the multivariate equivalent of repeated measures or mixed

ANOVA to assess variability within and between groups. By using this mixed ANOVA,

the separate anastomosis trials (RQ2) allowed us to examine the correlation between

separate data points and the rate of overall performance improvement experienced by the

participants. Comparison in time and leak rates with previous anastomosis labs (RQ4)

was evaluated using a Wilcoxon signed-rank test, as the data did not conform to a normal

distribution. To assess internal consistency between blinded examiners' MOSAT

checklist and GRS score, the inter-rater agreement and reliability was examined by

obtaining the Kappa coefficient along with the Pearson correlation, Kendall's tau, and

Spearman's rho for comparison. The relationship between variables on RQ1B and RQ3 was nonlinear; therefore, comparison of resident's self-evaluation of technical skill scores with the scores provided by the expert evaluators was analyzed using a Spearman correlation coefficient. All statistical analyses for this study was conducted using IBM SPSS version 25 (released 2017, IBM Corp. Armonk, NY). Statistical significance was set at a p-value equal to or less than 0.05.

Chapter Three Summary

This chapter described the methodology and procedures employed in an effort to compare self-assessment video analysis against structured, individualized, video-based coaching as a way to improve technical skill acquisition in surgical residents participating in a vascular anastomosis simulation workshop. The research questions, research design, sample population, instrumentation and procedures employed in this research were all presented and described accordingly. Additionally, this chapter discussed the data collection process this research utilized, as well as the approach selected for data analysis based on the information attained. The chapter concluded with some of the limitations associated with the project as a whole. The presentation of this data in chapter four will specifically address the research questions proposed, and the general demographic information of the research participants.

Chapter Four: Presentation and Analysis of the Data

Surgical training has changed dramatically over the past two decades in light of concerns over patient safety, resident oversight, and resident well-being. The cumulative effects of these changes, however, surrounding the competency of general surgeons after five years of residency, in this new training paradigm, have recently come to light. Because of this, there is a greater emphasis on the role of simulation lab training during surgical residency to improve technical skill acquisition, in an effort to efficiently achieve competency in designated core procedures. The purpose of this study was to compare self-assessment video analysis against structured, individualized, video-based coaching as a way to improve technical skill acquisition in novice, junior surgical residents learning how to perform a vascular anastomosis. The results from this research will now be presented and discussed in relation to each of the research questions.

Resident Participation

A total of thirteen (n=13) surgical residents participated in this vascular anastomosis lab which ran for a total of seven weeks during the summer of 2019, from July 8th through August 23rd. As the start of a new academic year always commences in the final week of June, each of the residents participating in this research had just begun their second year of surgical residency. This is important to note, as none of the residents had any practical experience performing an end to side vascular anastomosis prior to this, which was the focus of this simulation lab. The majority of the participants (n=9) were made up of residents from the department of general surgery. There were an additional four residents from two other surgical subspecialties who participated in the lab, as this anastomosis skill is also pertinent to their professional development. Two resident

participants were from the department of urologic surgery (n=2), and the remaining two

residents were from the department of plastic surgery (n=2). According to the 2010

census, women currently make up approximately 50% of graduating medical school

students but only 36% of surgical residents and 15% of active general surgeons (Bruce,

Battista, Plankey, Johnson, & Marshall, 2015). The distribution of female resident

participants in this research was slightly better than this national statistic, with a 46%

representation. In 2015, Black Americans represented only 5.7% of graduating medical

students and 6.2% of general surgery residents (Abelson et al., 2018). Resident

participants in this research mimicked this statistic, accounting for only 7% of the total

group. Table 1 provides an overview of the thirteen residents who participated.

Table 1: *Participant Attributes*

Resident	Gender	Coaching versus Non-Coaching	Hand Dominance	Specialty	Age
001	Male	Non-Coaching	Right	General Surgery	28
002	Female	Non-Coaching	Right	General Surgery	27
003	Male	Coaching	Right	General Surgery	32
004	Male	Coaching	Right	General Surgery	28
005	Female	Coaching	Right	General Surgery	27
006	Male	Coaching	Right	Urology	27
007	Female	Non-Coaching	Right	Urology	28
008	Female	Non-Coaching	Right	General Surgery	49
009	Female	Non-Coaching	Right	General Surgery	27
010	Male	Coaching	Right	General Surgery	27
011	Female	Non-Coaching	Right	General Surgery	27
012	Male	Non-Coaching	Right	Plastic Surgery	32
013	Male	Coaching	Right	Plastic Surgery	34

Inter-Rater Agreement and Reliability

In order to obtain a benchmark score for technical skill acquisition, two blinded faculty members scored each resident independently using the MOSAT and GRS evaluation tool. In order to obtain a single benchmark score for each resident trial, based off of these two data sets, the inter-rater agreement and inter-rater reliability of the scores obtained from the two blinded faculty members was calculated using Cohens Kappa statistic as well as the Pearson correlation, Kendall's tau, and Spearman's rho. Cohens Kappa statistic revealed a poor level of inter-rater agreement between the two blinded evaluators with the MOSAT k = 0.12, and GRS k of 0.01

Further analysis of the blinded faculty evaluators MOSAT and GRS data sets revealed a high level of concordance in both attending's scoring, however. All three correlation coefficients comparing the faculty MOSAT scoring, including Pearson's (0.86), Kendall's tau (0.70), and Spearman's rho (0.84) revealed a strong correlation. The same correlation was found between the faculty GRS scoring, with Pearson's (0.80), Kendall's tau (0.66), and Spearman's rho (0.83) all signifying that a strong level of concordance between the scores administered by the two faculty evaluators. Given these findings, the two sets of blinded faculty scores (attending 1 and attending 2) from the MOSAT and the GRS evaluation for each resident, were averaged to provide one single score. This single MOSAT and GRS score was then used as the benchmark score against which each resident would be compared, and would also compare themselves to, as part of their self-assessment and final analysis. See Table 2 for inter-rater agreement and reliability.

Table 2: *Inter-Rater Agreement and Reliability*

Attending 1 vs Attending 2	Kappa	Pearson Correlation	Kendall's tau	Spearman's rho
MOSAT	0.12	0.86	0.70	0.84
GRS	0.01	0.80	0.66	0.83

Baseline Performance (Pre-Test)

Prior to answering the research questions that guided this study, it is important to consider the baseline performance and ability of residents who participated in this anastomosis lab. Prior to the Pre-Test (week 1), the residents were randomly assigned to either the self-assessment video analysis only group (Non-Coaching) or the structured, individualized, video-based coaching group (Coaching). Participants arrived at the anastomosis lab in week one with no formal instruction and were asked to perform an end to side anastomosis per the protocol. The following Table 3 shows the Means and Standard Deviations for each measured variable after the Pretest performance.

Table 3: Group Means and Standard Deviations Week 1 (Pretest)

	Coaching vs Non-Coaching	N	Mean	Std. Deviation
Attending MOSAT	Coaching	6	9.42	2.11
	Non-Coaching	7	12.43	3.08
Resident MOSAT	Coaching	6	10.00	4.00
	Non-Coaching	7	13.29	2.69
Attending GRS	Coaching	6	14.25	2.07
	Non-Coaching	7	17.29	2.69
Resident GRS	Coaching	6	14.33	4.76
	Non-Coaching	7	17.14	2.91
Averaged Time	Coaching	6	24.17	3.31
	Non-Coaching	7	27.14	5.15
Leak Rate	Coaching	6	76.67	17.54
	Non-Coaching	7	51.86	18.89
Combined Technical Score	Coaching	6	-34.67	24.87
	Non-Coaching	7	5.29	25.01

To evaluate if there were any significant differences between the coaching and the noncoaching groups during the pretest week, an Independent samples T Test was performed.

With the exception of Resident GRS, there were no outliers in the data, as assessed by inspection of a boxplot. Reanalysis of the data for resident GRS, minus this one outlier, did not alter the significance for this category, therefore the outlier was maintained in the final analysis. All seven variables were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). Despite randomization of participants, there were significant differences in several categorical variables between the Coaching and Non-Coaching group for week one (Pre-Test). There was a difference in attending GRS scores between the Coaching ($M = 14.25$, $SD = 2.07$) and the Non-Coaching group ($M = 17.29$, $SD = 2.69$), with a statistically significance mean difference of 3.04, 95% CI [-6.01, -.06], $t(11) = -2.25$, $p = .046$. There was also a difference in the Leak Rate between the Coaching ($M = 76.67$, $SD = 17.54$) and the Non-Coaching group ($M = 51.86$, $SD = 18.89$), with a statistically significant mean difference of 24.81, 95% CI [2.42, 47.20], $t(11) = 2.44$, $p = .033$. Lastly, a difference was observed for the Combined Technical Score between the coaching ($M = -34.67$, $SD = 24.87$) and Non-Coaching group ($M = 5.29$, $SD = 25.01$), with a statistically significant mean difference of -29.38, 95% CI [-70.50, -9.40], $t(11) = -2.88$, $p = .015$. With respect to the other four variables measured, there were no significant differences observed in attending MOSAT score, resident self-evaluation MOSAT and GRS scores, or average time to complete anastomosis between the groups for week one (Pre-Test). See Table 4 and Graph 1 for a detailed overview of the categorical effects between groups and representation of this data.

Table 4: Week One Pre-Test Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Attending MOSAT Week 1	Equal variances assumed	.886	.367	-2.021	11	.068	-3.0119	1.490	-6.2919	.2681
Resident MOSAT Week 1	Equal variances assumed	.812	.387	-1.763	11	.106	-3.286	1.864	-7.388	.816
Attending GRS Week 1	Equal variances assumed	.423	.529	-2.248	11	.046	-3.0357	1.350	-6.0078	-.0636
Resident GRS Week 1	Equal variances assumed	1.010	.336	-1.307	11	.218	-2.810	2.149	-7.540	1.921
Averaged time Week1	Equal variances assumed	3.222	.100	-1.214	11	.250	-2.976	2.452	-8.373	2.421
Leak Rate Week 1	Equal variances assumed	.000	.985	2.439	11	.033	24.810	10.17	2.418	47.201
CTS Week 1	Equal variances assumed	.001	.976	-2.878	11	.015	-39.952	13.88	-70.502	-9.402

Graph 1: *Week One Pre-test Means*
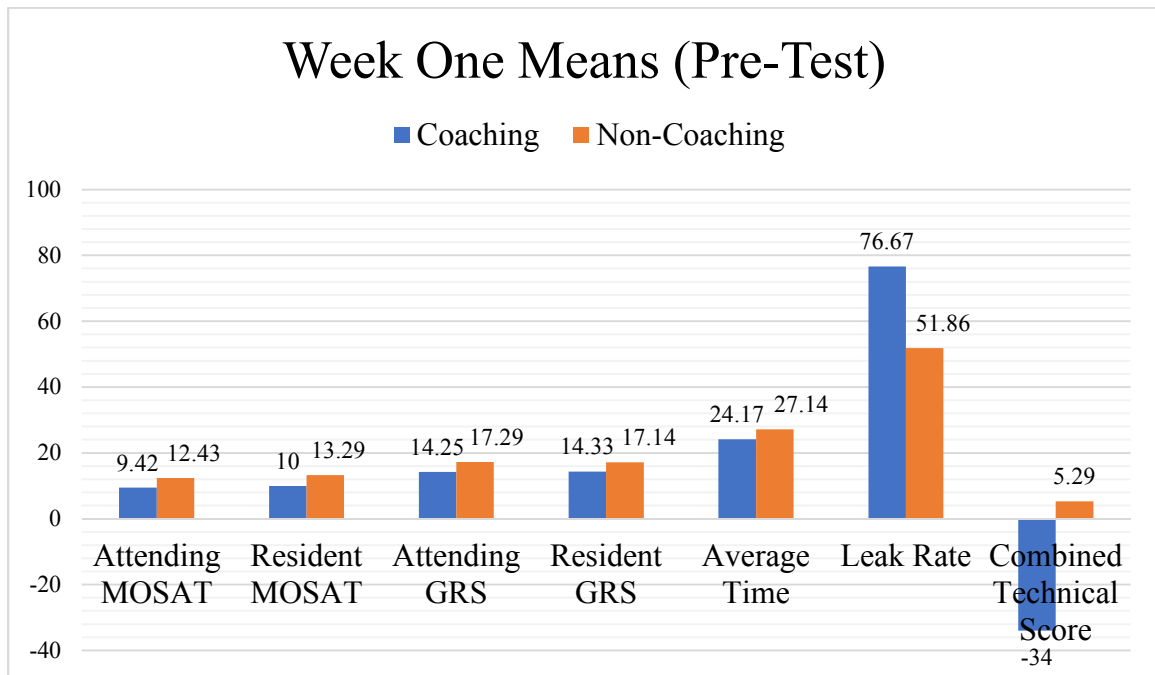
Graph 1: Week One Pre-test Means

RQ1A

Based on expert evaluator scoring using a validated technical skills checklist, to what extent do residents who participate in structured, individualized, video-based coaching attain technical skill, compared to residents in a self-assessment, video analysis only group?

In order to evaluate the effect coaching may have had on the technical skill acquisition of residents performing a vascular anastomosis, inferential statistics using a two-way repeated measures ANOVA with between groups design were run for Attending MOSAT, Attending GRS, Average Time, Leak Rate, and Combined Technical Score. All five of these variables under investigation were normally distributed ($p > .05$) across all four trials, except for Attending MOSAT week 5 ($p = .016$), as assessed by Shapiro-Wilk's test of normality on the studentized residuals. As the ANOVA is considered to be fairly robust with respect to deviations from normality, the PI chose not to transform the data from week five as it likely would not affect the overall results, which are presented next.

MOSAT

With regards to the attending MOSAT scores, there were two outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. These outliers occurred only in the experimental group (Coaching), once in week five and once in week seven (Post-Test). Removing these outliers with reanalysis of the data did not change the overall results, so they were both maintained for final analysis. With the exception of week 5 data, the data from week one (Pre-Test), week three, and week seven (Post-Test) was normally distributed, as assessed by Shapiro-

Wilk's test of normality ($p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 5.62$, $p = .347$.

The main effect of Attending MOSAT score showed that there was a statistically significant difference in scores within subjects across three of the four different time points for all participants, $F(3, 33) = 90.88$, $p < .001$, partial $\eta^2 = .89$. There was an increase in Attending MOSAT scores week one Pre-Test, ($M = 11.04$, $SD = 3.00$) to Week 7 Post-Test ($M = 20.15$, $SD = .55$), with a statistically significant mean increase of 9.23, 95% CI [11.33, 7.12], $p < .001$. The between-subjects effect of Coaching versus Non-Coaching did not show a significant difference in Attending MOSAT scores, $F(1, 11) = 2.78$, $p = .124$, partial $\eta^2 = .202$. See Table 5 and Graph 2 for descriptive and inferential statistical representation of this data. Refer to Table 6 for the mean differences in MOSAT scores compared by week.

Table 5: *Attending MOSAT Descriptive Statistics*

	Coaching or Non-Coaching	Mean	Std. Deviation	N
Attending MOSAT week 1	Coaching	9.417	2.1075	6
	Non-Coaching	12.429	3.0745	7
	Total	11.038	3.0032	13
Attending MOSAT week 3	Coaching	14.833	1.7795	6
	Non-Coaching	15.857	1.5999	7
	Total	15.385	1.6975	13
Attending MOSAT week 5	Coaching	18.917	2.2895	6
	Non-Coaching	19.286	1.1495	7
	Total	19.115	1.6975	13
Attending MOSAT week 7	Coaching	20.083	.6646	6
	Non-Coaching	20.214	.4880	7
	Total	20.154	.5547	13

Graph 2: *Attending MOSAT Score (Mean)*
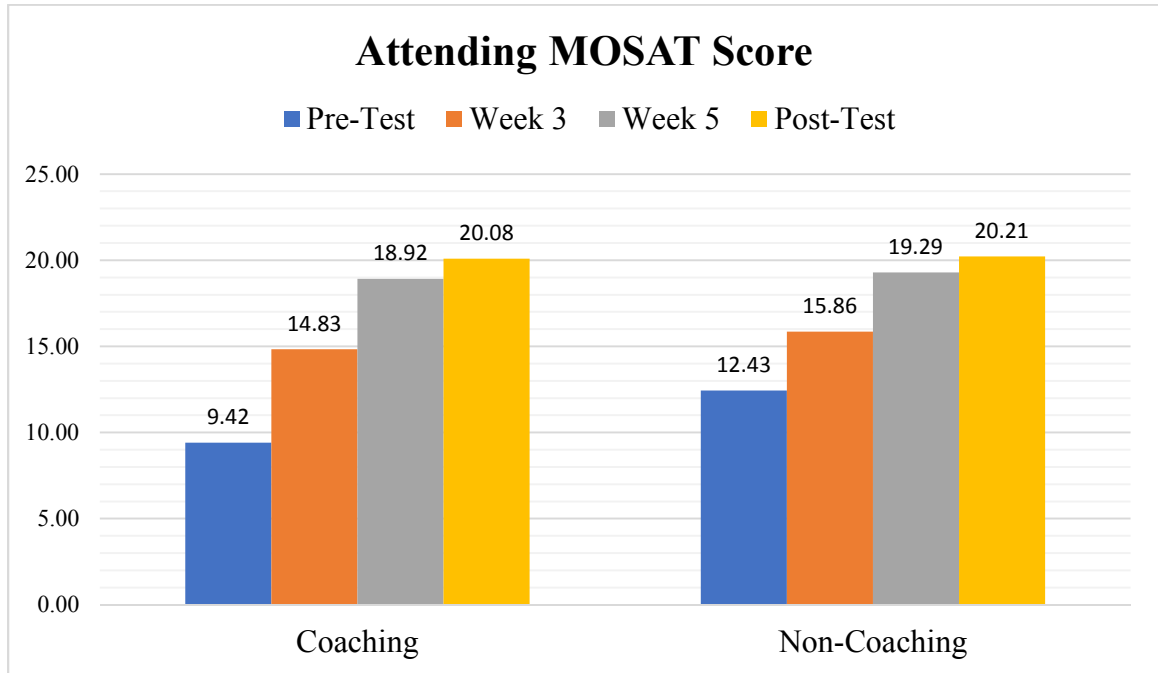
Table 6: *Pairwise Comparison Attending MOSAT Score*

(I) AttendMosat	(J) AttendMosat	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-4.423[*]	.701	.000	-6.673	-2.173
	3	-8.179[*]	.806	.000	-10.764	-5.593
	4	-9.226[*]	.656	.000	-11.332	-7.120
2	1	4.423[*]	.701	.000	2.173	6.673
	3	-3.756[*]	.559	.000	-5.549	-1.963
	4	-4.804[*]	.391	.000	-6.058	-3.549
3	1	8.179[*]	.806	.000	5.593	10.764
	2	3.756[*]	.559	.000	1.963	5.549
	4	-1.048	.521	.416	-2.718	.623
4	1	9.226[*]	.656	.000	7.120	11.332
	2	4.804[*]	.391	.000	3.549	6.058
	3	1.048	.521	.416	-.623	2.718

GRS

With regards to the Attending GRS scores, there were two outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. These outliers occurred only in the Non-Coaching group during week five and removing them did not change the overall results so both outliers were maintained for final analysis. The data from all four trial weeks were normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 2.67$, $p = .752$.

The main effect of Attending GRS score showed that there was a statistically significant difference in scores within subjects at all four different time points for all participants, $F(3, 33) = 208.63$, $p < .001$, partial $\eta^2 = .95$. There was an increase in Attending GRS scores week one Pre-Test, ($M = 15.88$, $SD = 2.81$) to Week 7 Post-Test ($M = 30.31$, $SD = 1.52$), with a statistically significant mean increase of 14.50, 95% CI [16.48, 12.52], $p < .001$. The between-subjects effect of Coaching versus Non-Coaching did not show a significant difference in Attending GRS scores, $F(1, 11) = .904$, $p = .362$, partial $\eta^2 = .076$. See table 7 and Graph 3 for descriptive and inferential statistical representation of this data. Refer to Table 8 for the mean differences in MOSAT scores compared by week.

Table 7: *Attending GRS Descriptive Statistics*

	Coaching or Non-Coaching	Mean	Std. Deviation	N
Attending GRS week 1	Coaching	14.250	2.0676	6
	Non-Coaching	17.286	2.6904	7
	Total	15.885	2.8074	13
Attending GRS week 3	Coaching	20.917	1.9343	6
	Non-Coaching	20.857	2.3223	7
	Total	20.885	2.0631	13
Attending GRS week 5	Coaching	26.167	2.6583	6
	Non-Coaching	25.857	2.5774	7
	Total	26.000	2.5083	13
Attending GRS week 7	Coaching	29.750	1.6355	6
	Non-Coaching	30.786	1.3496	7
	Total	30.308	1.5212	13

Graph 3: Attending GRS Score (Mean)
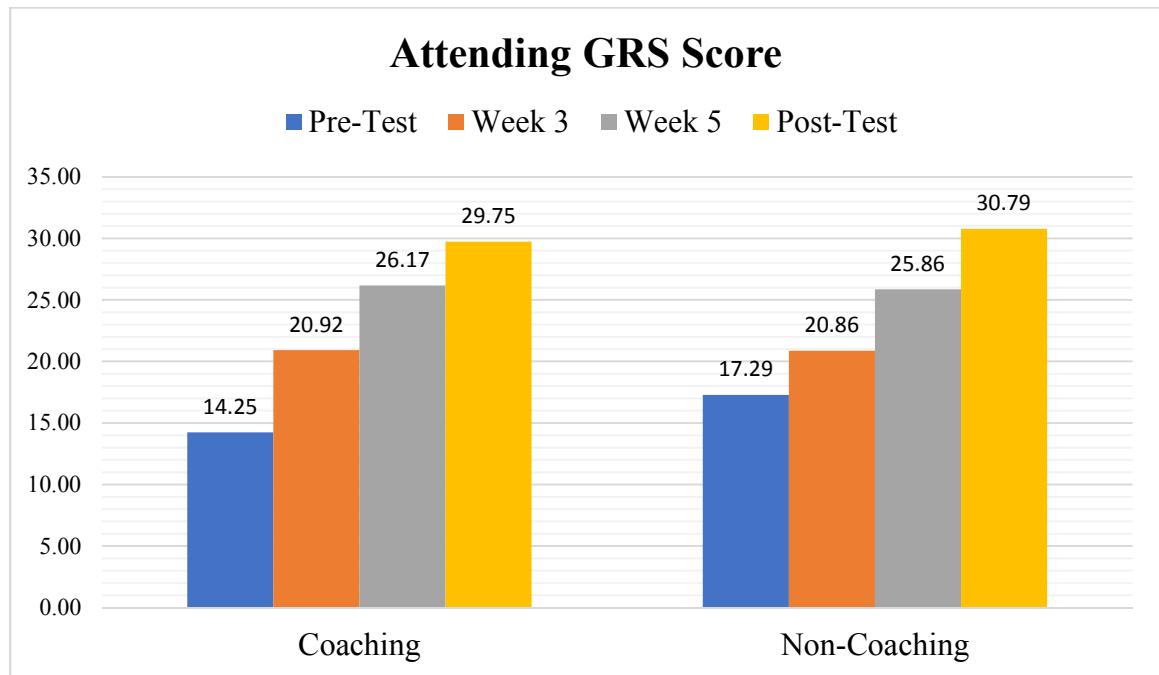
Table 8: *Pairwise Comparisons Attending GRS Score*

(I) AttendGRS	(J) AttendGRS	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-5.119[*]	.661	.000	-7.239	-3.000
	3	-10.244[*]	.619	.000	-12.229	-8.259
	4	-14.500[*]	.616	.000	-16.477	-12.523
2	1	5.119[*]	.661	.000	3.000	7.239
	3	-5.125[*]	.699	.000	-7.366	-2.884
	4	-9.381[*]	.452	.000	-10.831	-7.931
3	1	10.244[*]	.619	.000	8.259	12.229
	2	5.125[*]	.699	.000	2.884	7.366
	4	-4.256[*]	.616	.000	-6.232	-2.280
4	1	14.500[*]	.616	.000	12.523	16.477
	2	9.381[*]	.452	.000	7.931	10.831
	3	4.256[*]	.616	.000	2.280	6.232

Time

With regards to the time to complete the end-to-side anastomosis, there were no outliers in this data set, as assessed by inspection of a boxplot. The data from all four trial weeks were normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 7.54$, $p = .185$.

The main effect of time to complete the anastomosis showed that there was a statistically significant difference with an increase in time to complete the anastomosis within subjects, $F (3, 33) = 3.75$, $p < .001$, partial $\eta^2 = .25$. Pairwise analysis (Table 10) shows this difference was between week one (Pre-Test) and Week 3 (trial 2) only.

Overall, there was an increase in time to complete the anastomosis from week one Pre-Test, ($M = 25.77$, $SD = 4.49$) to Week 7 Post-Test ($M = 27.69$, $SD = 3.75$), with no statistically significant mean difference of 2.19, 95% CI [6.47, 2.09], $p = .772$. The between-subjects effect of Coaching versus Non-Coaching did not show a significant difference in time to complete the anastomosis, $F (1, 11) = .007$, $p = .934$, partial $\eta^2 = .001$. See Table 9 and Graph 4 for descriptive and inferential statistical representation of this data.

Table 9: *Averaged Time to Complete Anastomosis Descriptive Statistics*

	Coaching or Non-Coaching	Mean	Std. Deviation	N
Averaged time Week 1	Coaching	24.17	3.312	6
	Non-Coaching	27.14	5.146	7
	Total	25.77	4.494	13
Averaged time Week 3	Coaching	30.33	5.125	6
	Non-Coaching	29.29	7.825	7
	Total	29.77	6.470	13
Averaged time Week 5	Coaching	27.00	2.530	6
	Non-Coaching	29.86	6.817	7
	Total	28.54	5.301	13
Averaged time Week 7	Coaching	29.83	3.061	6
	Non-Coaching	25.86	3.436	7
	Total	27.69	3.750	13

Graph 4: *Average Time to Complete Anastomosis (Mean)*
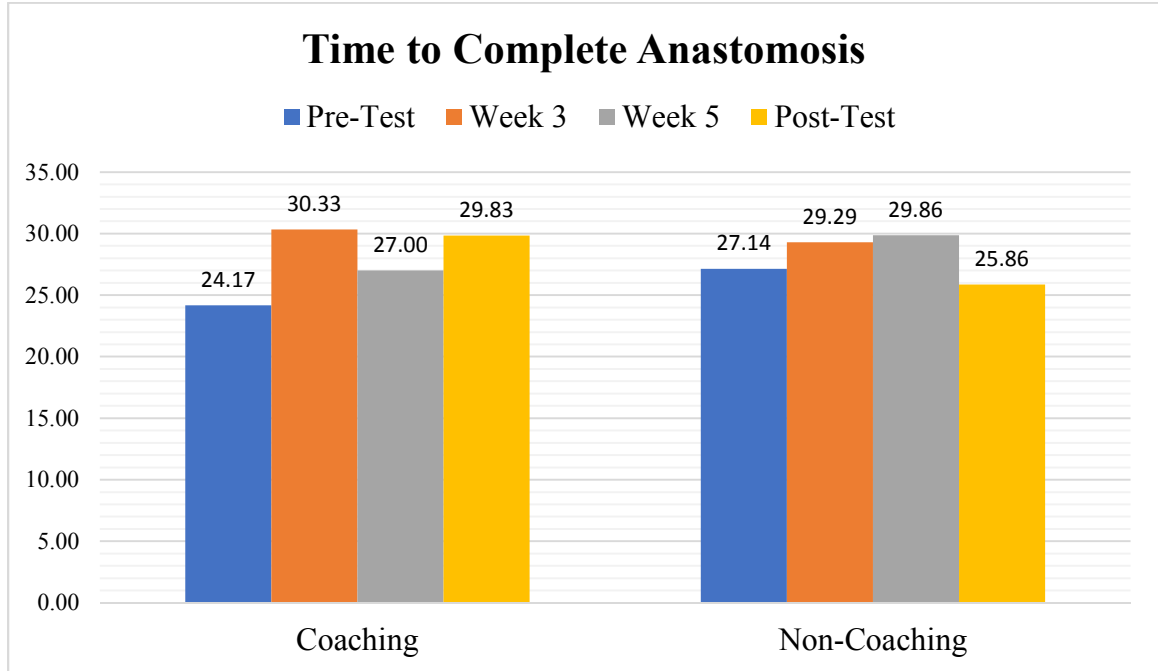
Table 10: *Averaged Time to Complete Anastomosis Pairwise Comparisons*

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-4.155*	1.019	.011	-7.425	-.885
	3	-2.774	1.068	.149	-6.201	.654
	4	-2.190	1.333	.772	-6.468	2.087
2	1	4.155*	1.019	.011	.885	7.425
	3	1.381	.960	1.000	-1.699	4.461
	4	1.964	1.730	1.000	-3.584	7.513
3	1	2.774	1.068	.149	-.654	6.201
	2	-1.381	.960	1.000	-4.461	1.699
	4	.583	1.296	1.000	-3.575	4.742
4	1	2.190	1.333	.772	-2.087	6.468
	2	-1.964	1.730	1.000	-7.513	3.584
	3	-.583	1.296	1.000	-4.742	3.575

Anastomotic Leak Rate

With regards to the amount of saline that ultimately leaked from the completed

anastomosis, there were four outliers in this data, as assessed by inspection of a boxplot

for values greater than 1.5 box-lengths from the edge of the box. These outliers occurred in two participants in the Coaching group and two participants in the Non-Coaching group during weeks three (n=1), five (n=3) and seven (Post-Test) (n=1). Because of the variability in average leak rate as a marker of technical skill (to be further discussed in chapter 5), these outliers were maintained for final analysis. The data from all four trial weeks were normally distributed, as assessed by Shapiro-Wilk's test of normality ($p >$ .05). There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 5.45, p = .365$.

The main effect of leak rate from the completed anastomosis showed that there was a statistically significant difference in leak rates within subjects, $F(3, 33) = 6.23, p = .002$, partial $\eta^2 = .362$. There was an overall decrease in leak rate from week one (Pre-Test), ($M = 63.31, SD = 21.73$) to Week 7 Post-Test ($M = 48.46, SD = 8.77$), with no statistically significant mean difference of 15.69, 95% CI [.271, 31.65], $p = .055$. Of note, there was a statistically significant mean difference of 17.55, 95% CI [32.05, 3.05], $p = .015$ from week one (Pre-test) to week five (see pairwise analysis in Table 12 for greater detail). The between-subjects effect of Coaching versus Non-Coaching did not show a significant difference in anastomotic leak rate, $F(1, 11) = 2.84, p = .120$, partial $\eta^2 = .205$. See table 11 and Graph 5 for descriptive and inferential statistical representation of this data.

Table 11: *Anastomotic Leak Rate Descriptive Statistics*

	Coaching or Non-coaching	Mean	Std. Deviation	N
Leak Rate week 1	Coaching	76.67	17.535	6
	Non-Coaching	51.86	18.889	7
	Total	63.31	21.731	13
Leak Rate week 3	Coaching	65.83	13.273	6
	Non-Coaching	54.00	20.363	7
	Total	59.46	17.845	13
Leak Rate week 5	Coaching	44.00	5.692	6
	Non-Coaching	49.43	9.572	7
	Total	46.92	8.200	13
Leak Rate week 7	Coaching	50.00	10.450	6
	Non-Coaching	47.14	7.647	7
	Total	48.46	8.771	13

Graph 5: *Anastomotic Leak Rate (Mean)*
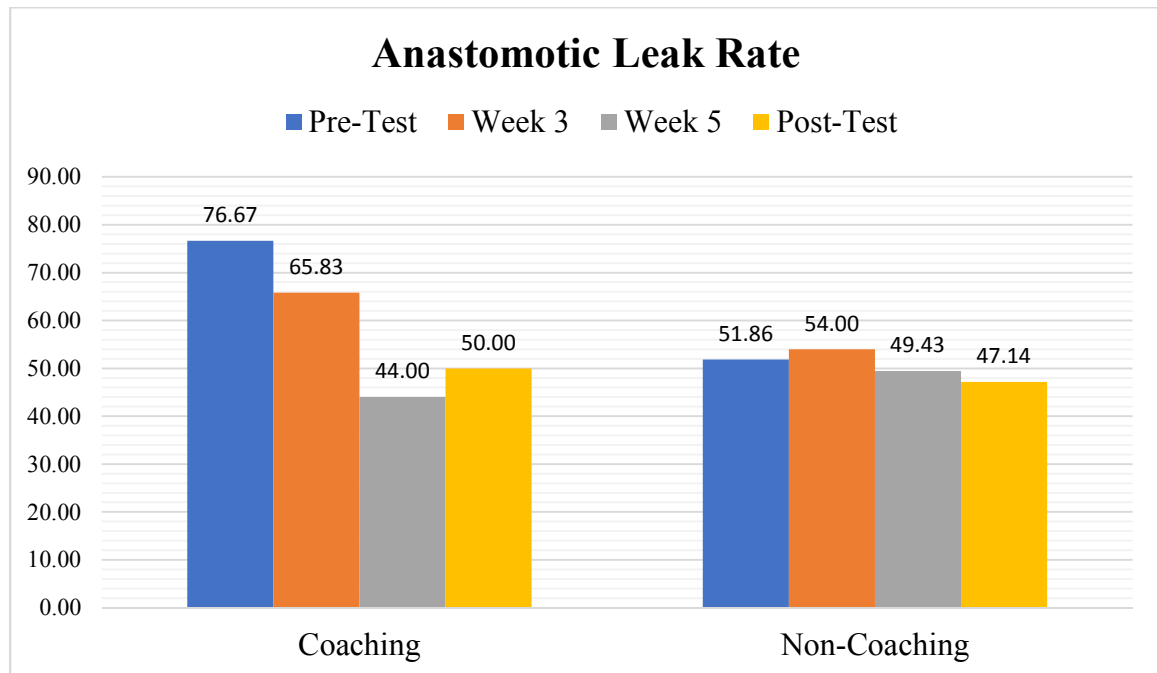
Table 12: *Leak Rate Pairwise Comparisons*

(I) Leak	(J) Leak	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	4.345	6.657	1.000	-17.011	25.702
	3	17.548*	4.521	.015	3.045	32.050
	4	15.690	4.975	.055	-.271	31.652
2	1	-4.345	6.657	1.000	-25.702	17.011
	3	13.202	4.448	.077	-1.068	27.472
	4	11.345	4.395	.153	-2.754	25.445
3	1	-17.548*	4.521	.015	-32.050	-3.045
	2	-13.202	4.448	.077	-27.472	1.068
	4	-1.857	3.550	1.000	-13.246	9.531
4	1	-15.690	4.975	.055	-31.652	.271
	2	-11.345	4.395	.153	-25.445	2.754
	3	1.857	3.550	1.000	-9.531	13.246

Combined Technical Score

In order to try and quantify a resident's technical skill acquisition, a combined technical score (CTS) was developed for this project using the objective measurements of MOSAT and GRS scores, time to complete the anastomosis, and leak rate. The formula upon which this CTS was calculated can be expressed by the following formula:

*Combined Technical Score = 2 * (MOSAT * 2) + (GRS * 2) – Time – Leak*

With regards to the resident's calculated CTS, there was one outlier in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. This outlier occurred in the Non-Coaching group in week seven (Post-Test). Because of the significance of this score, as a marker of technical skill, this outlier was removed from the data set and the data was re-run to assess for any change in significance. This, however, did not change the overall results, so this outlier was maintained in the final analysis.

The data from all four trial weeks were normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 9.99$, $p = .077$.

The main effect of CTS showed that there was a statistically significant difference in scores within subjects across the trials, $F(3, 33) = 74.45$, $p < .001$, partial $\eta^2 = .87$. There was an increase in CTS from the week one Pre-Test, ($M = -13.15$, $SD = 31.63$) to the Week 7 Post-Test ($M = 65.08$, $SD = 11.87$), with a statistically significant mean increase of 79.41, 95% CI [100.50, 58.31], $p < .001$. Pairwise analysis (Table 14) shows that significant differences in CTS occurred between weeks one, three, and five, but not between weeks five and seven. The between-subjects effect of Coaching versus Non-Coaching did not show a significant difference in Attending MOSAT scores, $F(1, 11) = 3.50$, $p = .088$, partial $\eta^2 = .241$. See Table 13 and Graph 6 for descriptive and inferential statistical representation of this data.

Table 13: *Combined Technical Score (CTS) Descriptive Statistics*

	Coaching or Non-Coaching	Mean	Std. Deviation	N
CTS Week 1	Coaching	-34.67	24.873	6
	Non-Coaching	5.29	25.011	7
	Total	-13.15	31.628	13
CTS Week 3	Coaching	5.00	11.933	6
	Non-Coaching	21.86	27.425	7
	Total	14.08	22.625	13
CTS Week 5	Coaching	57.00	18.612	6
	Non-Coaching	49.57	14.351	7
	Total	53.00	16.192	13
CTS Week 7	Coaching	60.00	10.677	6
	Non-Coaching	69.43	11.788	7
	Total	65.08	11.871	13

Graph 6: *Combined Technical Score (Mean)*
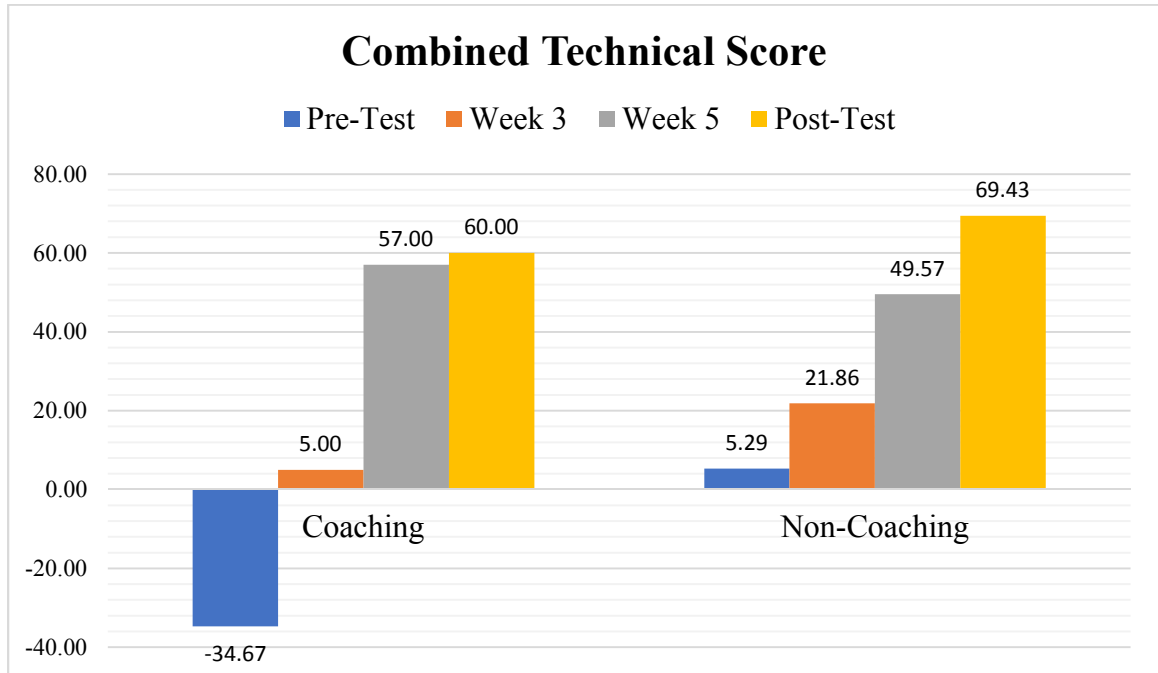
Table 14: *Combined Technical Score Pairwise Comparisons*

(I) NewScore	(J) NewScore	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-28.119*	8.134	.032	-54.214	-2.024
	3	-67.976*	4.587	.000	-82.691	-53.261
	4	-79.405*	6.575	.000	-100.500	-58.310
2	1	28.119*	8.134	.032	2.024	54.214
	3	-39.857*	6.702	.001	-61.359	-18.356
	4	-51.286*	4.659	.000	-66.233	-36.339
3	1	67.976*	4.587	.000	53.261	82.691
	2	39.857*	6.702	.001	18.356	61.359
	4	-11.429	4.328	.138	-25.314	2.456
4	1	79.405*	6.575	.000	58.310	100.500
	2	51.286*	4.659	.000	36.339	66.233
	3	11.429	4.328	.138	-2.456	25.314

RQ1B

Using the validated technical skills checklist, to what extent do resident's self-assessment scoring of their own technical skill differ between those who participate in the structured, individualized, video-based coaching group, compared to those in the self-assessment, video analysis only group?

In order to evaluate the effect coaching may have had on self-assessment scoring for residents performing this end-to-side vascular anastomosis, inferential statistics using a two-way repeated measures ANOVA, with between groups design, was performed.

Resident MOSAT

With regards to resident self-assessment MOSAT scores, there were no outliers in this data set, as assessed by inspection of a boxplot. The data were normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 2.23$, $p = .817$.

The main effect of resident self-assessment scores showed that there was a statistically significant difference in MOSAT scores within groups, $F(3, 33) = 69.82$, $p < .001$, partial $\eta^2 = .864$. There was an overall increase in Resident self-assessment MOSAT scores from week one Pre-Test, ($M = 11.77$, $SD = 3.63$) to Week 7 Post-Test ($M = 19.77$, $SD = 2.98$), with a statistically significant mean difference of 8.04, 95% CI [10.37, 5.70], $p < .001$. Pairwise analysis (Table 16) shows that significant differences in Resident MOSAT scores occurred between weeks one, three, and five, but not between

weeks five and seven. The between-subjects effect of Coaching versus Non-Coaching

did not show a significant difference in resident self-assessment MOSAT scores, F (1,

11) = 2.35, p = .154, partial η^2 = .176. See Table15 and Graph 7 for descriptive and

inferential statistical representation of this data.

Table 15: *Resident Self-Assessment MOSAT Score Descriptive Statistics*

	Coaching/Non-Coaching	Mean	Std. Deviation	N
Resident MOSAT week 1	Coaching	10.00	4.000	6
	Non-Coaching	13.29	2.690	7
	Total	11.77	3.632	13
Resident MOSAT week 3	Coaching	15.67	3.266	6
	Non-Coaching	17.71	2.059	7
	Total	16.77	2.774	13
Resident MOSAT week 5	Coaching	19.00	2.366	6
	Non-Coaching	19.71	2.360	7
	Total	19.38	2.293	13
Resident MOSAT week 7	Coaching	18.50	3.937	6
	Non-Coaching	20.86	1.345	7
	Total	19.77	2.976	13

Graph 7: *Resident Self-Assessment MOSAT Scores (Mean)*

86
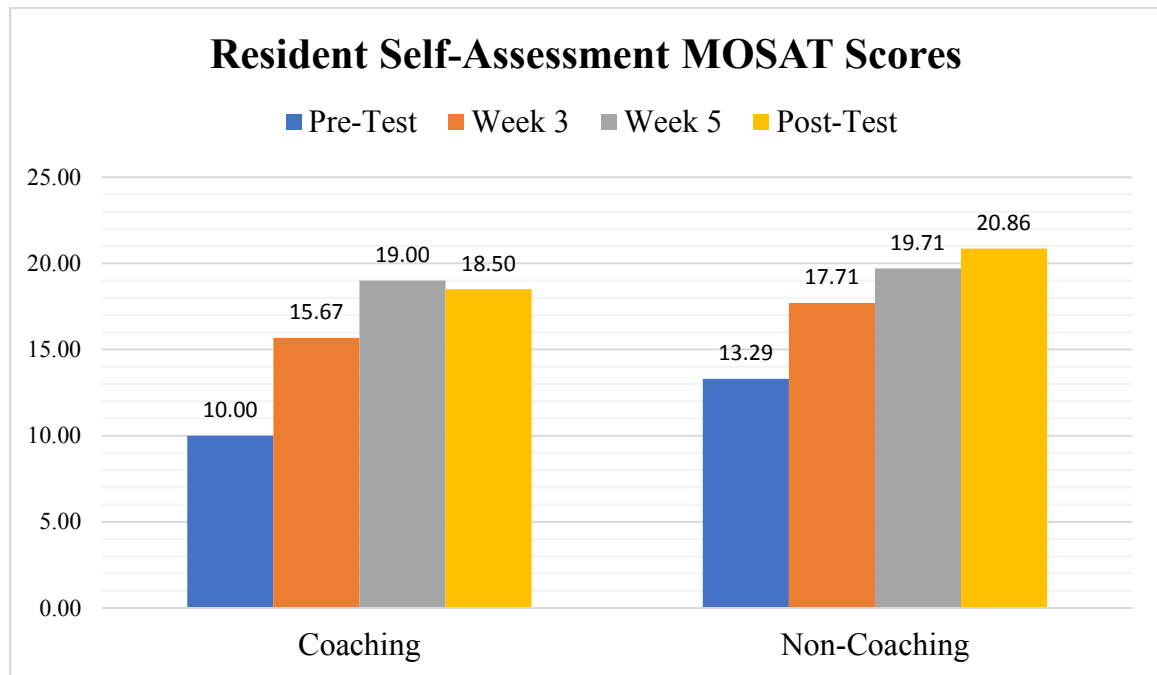
Table 16: *Pairwise Comparisons Resident MOSAT*

(I) ResidentMOSAT	(J) ResidentMOSAT	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-5.048[*]	.655	.000	-7.150	-2.945
	3	-7.714[*]	.637	.000	-9.758	-5.670
	4	-8.036[*]	.727	.000	-10.369	-5.702
2	1	5.048[*]	.655	.000	2.945	7.150
	3	-2.667[*]	.565	.004	-4.478	-.855
	4	-2.988[*]	.669	.006	-5.135	-.841
3	1	7.714[*]	.637	.000	5.670	9.758
	2	2.667[*]	.565	.004	.855	4.478
	4	-.321	.492	1.000	-1.899	1.256
4	1	8.036[*]	.727	.000	5.702	10.369
	2	2.988[*]	.669	.006	.841	5.135
	3	.321	.492	1.000	-1.256	1.899

Based on estimated marginal means
*. The mean difference is significant at the
b. Adjustment for multiple comparisons: Bonferroni.

Resident GRS

With regards to resident self-assessment GRS scores, there were two outliers in the Non-Coaching group. One outlier was in week one (Pre-Test), and the next was in week five. Removing these outliers did not alter the significance of the variable, therefore, they were maintained for final analysis to maintain consistency. The data were normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(5) = 3.99$, $p = .553$.

The main effect of resident self-assessment scores showed that there was a statistically significant difference in GRS scores within subjects, $F(3, 33) = 36.48$, $p <$

.001, partial η^2 = .768. There was an overall increase in Resident self-assessment GRS scores from week one Pre-Test, (M = 15.85, SD = 3.98) to Week 7 Post-Test (M = 28.23, SD = 6.87), with a statistically significant mean difference of 12.43, 95% CI [17.78, 7.08], p < .001. Pairwise analysis (Table 18) shows that significant differences in Resident GRS scores occurred between weeks one, three, and five, but not between weeks five and seven. The between-subjects effect of Coaching versus Non-Coaching did not show a significant difference in resident self-assessment GRS scores, F (1, 11) = .273, p = .612, partial η^2 = .024. See Table 17 and Graph 8 for descriptive and inferential statistical representation of this data.

Table 17: *Resident Self-Assessment GRS Score Descriptive Statistics*

	Coaching or Non-Coaching	Mean	Std. Deviation	N
Resident GRS week 1	Coaching	14.33	4.761	6
	Non-Coaching	17.14	2.911	7
	Total	15.85	3.976	13
Resident GRS week 3	Coaching	22.33	3.670	6
	Non-Coaching	24.29	5.619	7
	Total	23.38	4.735	13
Resident GRS week 5	Coaching	27.33	4.412	6
	Non-Coaching	26.00	4.619	7
	Total	26.62	4.388	13
Resident GRS week 7	Coaching	27.33	7.230	6
	Non-Coaching	29.00	7.024	7
	Total	28.23	6.870	13

Graph 8: *Resident Self-Assessment GRS Scores (Mean)*

Resident Self-Assessment GRS Scores

- Pre-Test
- Week 3
- Week 5
- Post-Test

Coaching: Pre-Test 14.33, Week 3 22.33, Week 5 27.33, Post-Test 27.33

Non-Coaching: Pre-Test 17.14, Week 3 24.29, Week 5 26.00, Post-Test 29.00
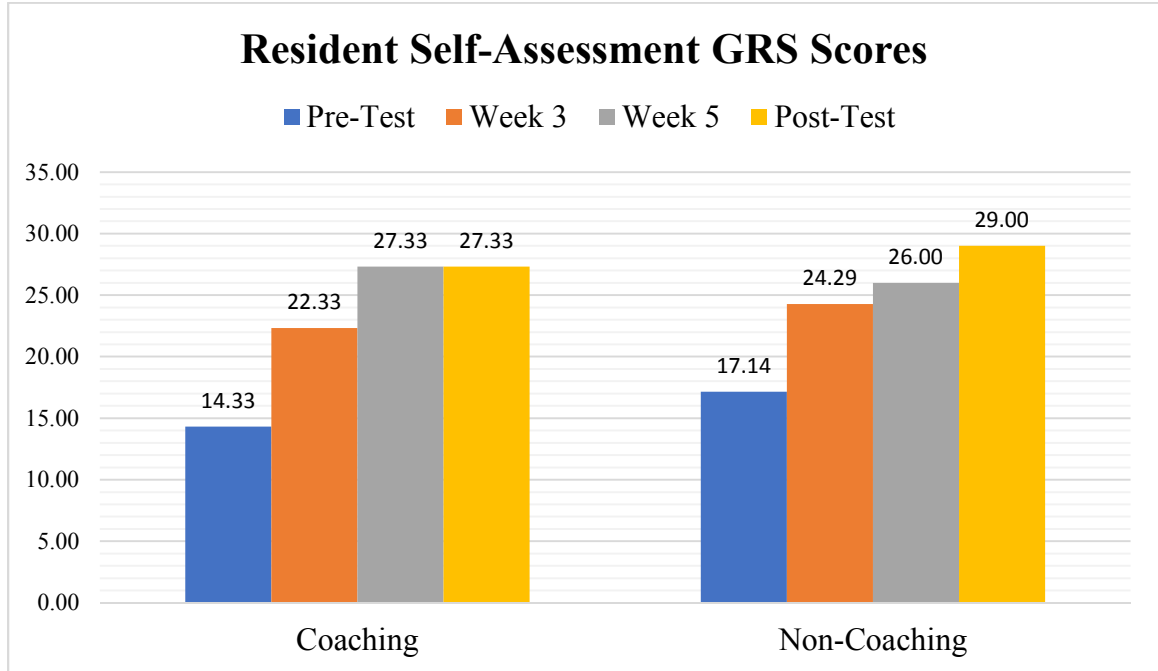
Table 18: *Pairwise Comparisons Resident GRS*

(I) ResidentGRS	(J) ResidentGRS	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-7.571*	1.285	.001	-11.695	-3.448
	3	-10.929*	1.134	.000	-14.567	-7.290
	4	-12.429*	1.667	.000	-17.775	-7.082
2	1	7.571*	1.285	.001	3.448	11.695
	3	-3.357*	.917	.022	-6.298	-.417
	4	-4.857*	1.287	.018	-8.986	-.729
3	1	10.929*	1.134	.000	7.290	14.567
	2	3.357*	.917	.022	.417	6.298
	4	-1.500	1.373	1.000	-5.905	2.905
4	1	12.429*	1.667	.000	7.082	17.775
	2	4.857*	1.287	.018	.729	8.986
	3	1.500	1.373	1.000	-2.905	5.905

Based on estimated marginal means

*. The mean difference is significant at the

b. Adjustment for multiple comparisons: Bonferroni.

RQ2

Based on expert evaluator scoring using the validated technical skills checklist, do residents who participate in structured, individualized, video-based coaching attain technical skill at a different rate than residents in the self-assessment video analysis only group?

A repeated measures ANOVA was conducted to determine if the rate of technical skill acquisition using the Attending MOSAT was different for residents participating in the Coaching versus the Non-Coaching group. Overall mean Attending MOSAT scores increased from the Pre-Test in week one ($M = 11.04$, $SD = 3.00$), to week three ($M = 15.39$, $SD = 1.70$), to week five ($M = 19.12$, $SD = 1.70$) to the Post-Test in week seven ($M = 20.15$, $SD = .55$) in that order. The differences between the coaching group and the non-coaching group, however, was not statistically significant at any point in time. See Table 19 for between group comparisons.

Table 19: *Attending MOSAT ANOVA*

		Sum of Squares	df	Mean Square	F	Sig.
Attending MOSAT week 1	Between Groups	29.308	1	29.308	4.085	.068
	Within Groups	78.923	11	7.175		
	Total	108.231	12			
Attending MOSAT week 3	Between Groups	3.386	1	3.386	1.194	.298
	Within Groups	31.190	11	2.835		
	Total	34.577	12			
Attending MOSAT week 5	Between Groups	.440	1	.440	.142	.714
	Within Groups	34.137	11	3.103		
	Total	34.577	12			
Attending MOSAT week 7	Between Groups	.055	1	.055	.168	.690
	Within Groups	3.637	11	.331		
	Total	3.692	12			

A one-way ANOVA was then repeated in order to determine if the rate of technical skill acquisition using the Attending GRS was different for residents participating in the Coaching versus the Non-Coaching group. Overall Attending GRS scores increased from the Pre-Test in week one ($M = 15.89$, $SD = 2.81$), to week three ($M = 20.89$, $SD = 2.06$), to week five ($M = 26.00$, $SD = 2.51$) to the Post-Test in week seven ($M = 30.30$, $SD = 1.52$) in that order. With the exception of the baseline skills assessment in the week one Pre-Test, the differences between the coaching group and the non-coaching group was not statistically significant with intervention at week three, five, or seven (Post-Test). See Table 20 for between group comparisons.

Table 20: *Attending GRS ANOVA*

		Sum of Squares	df	Mean Square	F	Sig.
Attending GRS week 1	Between Groups	29.773	1	29.773	5.054	.046
	Within Groups	64.804	11	5.891		
	Total	94.577	12			
Attending GRS week 3	Between Groups	.011	1	.011	.002	.961
	Within Groups	51.065	11	4.642		
	Total	51.077	12			
Attending GRS week 5	Between Groups	.310	1	.310	.045	.835
	Within Groups	75.190	11	6.835		
	Total	75.500	12			
Attending GRS week 7	Between Groups	3.466	1	3.466	1.569	.236
	Within Groups	24.304	11	2.209		
	Total	27.769	12			

RQ3

Based on the validated technical skills checklist, to what extent do resident self-assessment scores of their technical skill correlate with expert evaluator scores?

A Spearman's rank-order correlation was run to assess the relationship between resident self-assessment MOSAT scores and Attending MOSAT scores across the four separate trials. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot. There was no statistically significant correlation between resident self-assessment MOSAT scores and Attending MOSAT scores across four separate trials, $r_s(11) = .180$, $p = .556$, $r_s(11) = -.165$, $p = .591$, $r_s(11) = .478$, $p = .098$, , $r_s(11) = -.369$, $p = .215$, in that order. This correlation did not change significance when participants were isolated into Coaching and Non-Coaching groups and compared to attending scores. See Table 21 for resident and attending scoring for all four trials.

Table 21: *Resident and Attending MOSAT Scores and Correlation*

Resident	Week 1		Week 3		Week 5		Week 7	
	Attend MOSAT	Resident MOSAT	Attend MOSAT	Resident MOSAT	Attend MOSAT	Resident MOSAT	Attend MOSAT	Resident MOSAT
001	13	18	14	19	20.5	22	20	21
002	10.5	15	15.5	20	18.5	21	20	22
003	10	7	17	14	19	17	20.5	13
004	13	9	15.5	18	21	21	21	21
005	8.5	17	13.5	21	19	22	20	22
006	10	6	16.5	12	20	16	20	14
007	10.5	14	18	19	20.5	22	20.5	22
008	13.5	12	14.5	18	20.5	20	19.5	22
009	17.5	13	18	16	18.5	16	21	19
010	8	9	12.5	15	14.4	18	20	20
011	14	10	16	18	18	17	20.5	19
012	8	11	15	14	18.5	20	20	21
013	7	12	14	14	20	20	19	21
	$r_s = .180$		$r_s = -.165$		$r_s = .478$		$r_s = -.369$	

A Spearman's rank-order correlation was also run to assess the relationship between resident self-assessment GRS scores and Attending GRS scores across four separate trials. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot. There was no statistically significant correlation between resident self-assessment GRS scores and Attending GRS scores across four

separate trials, $r_s(11) = .402$, $p = .173$, $r_s(11) = -.235$, $p = .440$, , $r_s(11) = .448$, $p = .125$, , $r_s(11) = -.389$, $p = .188$, in that order. This correlation did not change significance when participants were isolated into Coaching and Non-Coaching groups and compared to attending scores. See Table 22 for resident and attending scoring for all four trials.

Table 22: *Resident and Attending GRS Scores and Correlation*

Resident	Week 1		Week 3		Week 5		Week 7	
	Attend GRS	Resident GRS	Attend GRS	Resident GRS	Attend GRS	Resident GRS	Attend GRS	Resident GRS
001	18.5	19	22	24	30.5	27	31.5	26
002	13	15	17.5	29	22	26	29.5	39
003	16	12	24	17	27	20	31.5	22
004	14.5	10	20	20	29.5	29	31	28
005	15	16	19.5	25	26	31	27.5	33
006	13.5	11	22.5	21	24.5	27	30	17
007	17.5	21	22.5	33	27	35	29.5	38
008	18.5	12	17.5	19	24.5	24	29.5	25
009	21	17	22.5	19	25.5	21	31.5	21
010	10.5	14	19	27	22	25	28	27
011	18	18	22.5	27	26	27	33	30
012	14.5	18	21.5	19	25.5	22	31	24
013	16	23	20.5	24	28	32	30.5	37
	$r_s = .402$		$r_s = -.235$		$r_s = .448$		$r_s = -.389$	

RQ4

To what extent do the current posttest scores on time to fashion anastomosis and leak rate compare with the prior three years of resident vascular anastomosis scores?

Prior vascular anastomosis labs at our current institution have followed a similar, albeit simplified, format in an effort to assess resident's technical skill. Using time to complete the anastomosis and the amount of saline leak from the completed anastomosis provided one measure to assess resident's baseline technical skill (Pre-Test) and compare it with any potential improvement in performance during the final week of the vascular

anastomosis lab (Post-Test). The score that had been used to assess differences in Pre-

Test and Post-Test was based upon the following formula:

$$Score = 300 - (Time \times 10) - Leak\ Rate$$

In an effort to assess how current results from the 2019 vascular anastomosis lab

held up to this simplified formula, we compared the prior three years of resident's scores

to what the current resident's scores would be with this simplified formula. Using this

simple score yielded the following results across years in Table 23 (note: each year

represents a different resident class):

Table 23: *2017-2019 Vascular Anastomosis Scores (Simplified Score)*

	2016		2017		2018		2019	
Resident #	Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test
1	64	125	64	160	142	76	139	92
2	137	115	82	129	-40	165	137	72
3	250	210	110	127	94	85	103	43
4	127	150	107	119	96	75	111	90
5	158	110	-2	75	136	86	141	18
6	78	105	31	106	-2	135	117	84
7	61	180	-32	53	44	-2	20	102
8	200	120	60	85	125	-13	49	92
9			-140	-5	26	36	-5	86
10							160	9
11							88	36
12							135	140
13							178	66

As some of the data was not normally distributed, a Wilcoxon signed-rank test

was conducted to determine the results of resident technical skill using the simplified

formula, as done in years past at the current institution. The data from resident

participants who participated in the vascular anastomosis lab in 2016, 2017, 2018, and

2019 (current participants) was evaluated. Both time to complete the anastomosis and

saline leak rate from the completed anastomosis was measured and compiled into the simplified formula (300 – (Time * 10 – Leak) pre vascular anastomosis workshop (week one) and during the final week of vascular anastomosis as a post-intervention.

The difference scores were approximately symmetrically distributed, as assessed by a histogram with superimposed normal curve. Data are medians unless otherwise stated.

For the 2016 cohort of resident participants, participation in the anastomosis lab, as graded by the simplified score, elicited an increase in simplified score for four participants post-intervention, whereas four participant's score decreased post-intervention. There was a median decrease in simplified score (*Mdn* 9.50) from pre-intervention (*Mdn* = 132) to post-intervention (*Mdn* = 122.50) and this difference was not statistically significant, $z = -.140$, $p = .889$.

For the 2017 cohort of resident participants, participation in the anastomosis lab, as graded by the simplified score, elicited an increase in simplified score for all nine participants post-intervention. There was a statistically significant median increase in simplified score (*Mdn* 46) from pre-intervention (*Mdn* = 60) to post-intervention (*Mdn* = 106), $z = -2.67$, $p = .008$.

For the 2018 cohort of resident participants, participation in the anastomosis lab, as graded by the simplified score, elicited an increase in simplified score for three participants post-intervention, whereas six participant's score decreased post-intervention. There was a median decrease in simplified score (*Mdn* 18) from pre-intervention (*Mdn* = 94) to post-intervention (*Mdn* = 76) and this difference was not statistically significant, $z = -.533$, $p = .594$.

For the 2019 cohort of resident participants, participation in the anastomosis lab, as evaluated by this simplified score, only elicited an increase in the simplified score for four participants post-intervention. The nine other 2019 participant's scores actually decreased post-intervention. There was a median decrease in simplified score (*Mdn* 33) from pre-intervention (*Mdn* = 117) to post-intervention (*Mdn* = 84) and this difference was not statistically significant, $z = -1.50$, $p = .133$. When this 2019 cohort was broken down into Coaching versus Noncoaching groups, the results did not change. In the coaching group, all six participant's simplified scores decreased post-intervention. There was a statistically significant median decrease in simplified scores (*Mdn* 74.5) from pre-intervention (*Mdn* = 129) to post-intervention (*Mdn* = 54.50), $z = -2.20$, $p = .028$. In the non-coaching group, three participant scores decreased, and four participant scores increased. There was a median increase in simplified scores (*Mdn* 4) from pre-intervention (*Mdn* = 88) to post-intervention (*Mdn* = 99) and this difference was not statistically significant, $z = -.338$, $p = .735$.

Refer to Table 24, Table 25, Table 26, Table 27, and Graph 9 for Z scores and medians for each cohort.

Table 24: *Resident Z Scores based on Simplified Score*

	Posttest 16 - Pretest 16	Posttest 17 - Pretest 17	Posttest 18 - Pretest 18	Posttest 19 - Pretest 19
Z	-.140[b]	-2.666[b]	-.533[c]	-1.503[c]
Asymp. Sig. (2-tailed)	.889	.008	.594	.133

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.
c. Based on positive ranks.

Table 25: *Simplified Median Scores*

Pretest 16	Posttest 16	Pretest 17	Posttest 17	Pretest 18	Posttest 18	Pretest 19	Posttest 19
132.00	122.50	60.00	106.00	94.00	76.00	117.00	84.00

Table 26: *2019 Z Scores (Coaching Vs Non-Coaching) Based on Simplified Score*

	Posttest 19 – Pretest 19	Coaching post 19 - Coaching pre 19	Noncoaching post19 - Noncoaching pre 19
Z	-1.503[b]	-2.201[b]	-.338[c]
Asymp. Sig. (2-tailed)	.133	.028	.735

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.
c. Based on negative ranks.

Table 27: *2019 Simplified Median Scores Coaching Vs Non-Coaching*

Pre-Test 2019	Posttest 2019	Coaching Pre 2019	Coaching Post 2019	Noncoaching Pre 19	Noncoaching Post 19
117.00	84.00	129.00	54.50	88.00	92.00

Graph 9: *2016-2019 Simplified Scores (Median)*
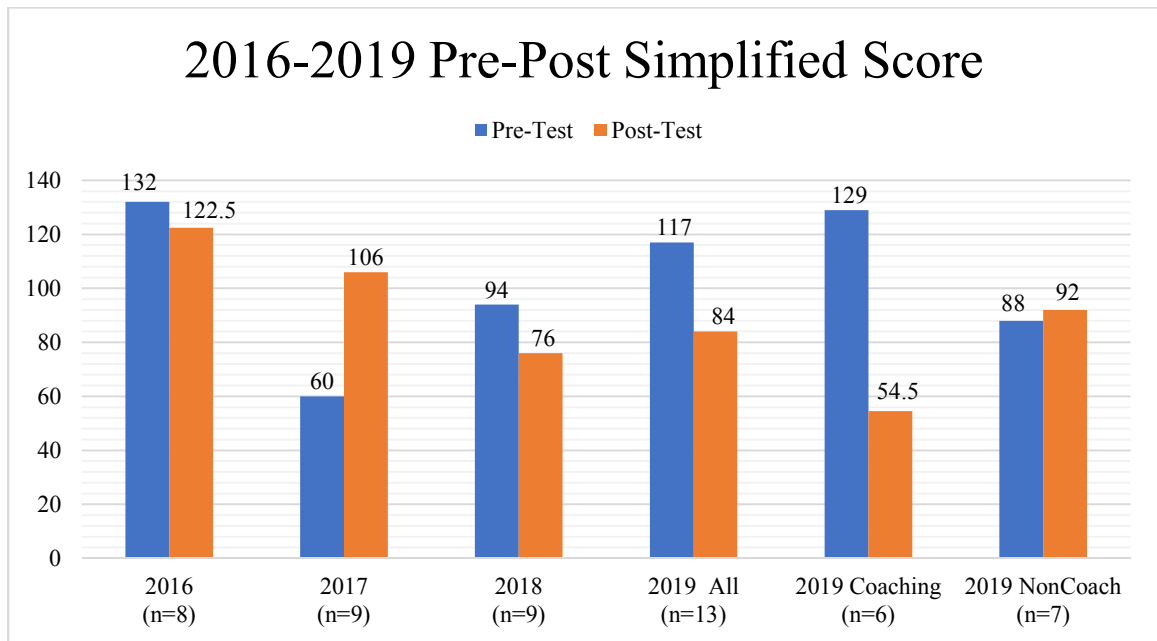
Chapter Four Summary

This chapter began with an overview of the purpose of this research and an overview of the characteristics of the thirteen residents who participated in this vascular anastomosis simulation lab. Variables of technical skill acquisition between the

structured, individualized, video-based coaching group and the self-assessment video

analysis only (Non-Coaching) group were then compared using descriptive and

inferential statistics. The data, as presented, did not show a statistical significance in

technical skill acquisition between the Coaching and Non-Coaching group. There were

significant improvements in the variables used to measure technical skill acquisition

between Pre-Test and Post-Test in both groups, however. These variables included

Attending MOSAT, Attending GRS, and Combined Technical Scores. Rates of

improvement in these variables between the Coaching and Non-Coaching groups did not

differ significantly across the seven weeks of the anastomosis lab.

With respect to resident self-assessment scoring, although there were statistically

significant mean increases in both Resident MOSAT and Resident GRS scores between

Pre-Test and Post-Test, the correlation coefficient between resident scores and attending

scores were found to be low to moderate across all four trials. With respect to assessing

technical competency using a simplified score based solely on time and leak rate, our

current vascular simulation lab participants (2019) did not show any significance in

improvement, which was similar to the 2016, and 2018 resident class. This also held true

when comparing Coaching versus Non-Coaching groups using this simplified score for

the 2019 residency class.

These results suggest there was no difference between structured, individualized,

video-based coaching and self-assessment video analysis only (Non-Coaching) as a

modality to improve technical skills in a vascular anastomosis simulation lab. The results

from this quantitative research will add to the literature on the use of coaching methods,

video analysis, and self-assessment of technical skill acquisition in surgical residency

programs. Chapter five will provide an interpretation of the data along with a detailed discussion based on the results. Findings from this research will be presented in a manner that extends the knowledge base and theoretical foundations contained in the accompanying literature review. In addition, limitations and suggestions for future research will be presented.

Chapter Five: Discussion, Conclusions, and Implications

The purpose of this research was to employ a validated technical skills checklist to compare structured, individualized, video-based coaching against self-assessment video analysis as a way to improve technical skill acquisition in surgical residents participating in a vascular anastomosis simulation workshop. This IRB approved research was conducted in a large academic, level one trauma center in the Northeastern United States with thirteen second-year junior surgical residents. While all participating residents significantly improved their technical skills related to fashioning an end-to-side vascular anastomosis, this research was unable to show a difference between the structured, individualized, video-based coaching group and the self-assessment only (Non-Coaching) group. A detailed discussion of these results, as they relate to each research question, will be presented next.

Discussion

The policy changes implemented in graduate medical education over the past two decades have forced surgical residency programs to become more efficient in their approach to educating residents. The shift away from a sole reliance on the apprenticeship model of training, and the movement towards the adoption and application of adult learning theory, has compelled surgical educators to rethink the way we teach basic technical skills to residents. In an effort to help residents attain technical competency, which will allow them to independently perform core surgical procedures by their fifth and final chief year of training, strategies such as video analysis and coaching models are continuously being evaluated. The research highlighted here was undertaken

as a way to incorporate a variety of theoretical frameworks and core concepts that support human psychomotor development in an effort to improve technical skill acquisition.

This study was designed to evaluate and compare the effect of structured, individualized, video-based coaching against self-assessment video analysis as a way to improve technical skill acquisition. Multiple studies have compared coaching models in surgery against a standard curriculum, or employed video-based analysis with and without feedback, as a methodology for teaching technical skills (Backstein et al., 2004, 2005; Farquharson, Cresswell, Beard, & Chan, 2013; Singh et al., 2015; Soucisse et al., 2017). To this author's knowledge, however, this is the first study to have investigated whether video-based coaching is superior to self-assessment video analysis. While the data, as presented here, was unable to show a significant difference between the two modalities, there are some fundamental concepts that need be acknowledged and discussed.

Careful attention to the structure and setup of this vascular anastomosis simulation lab was employed in a manner consistent with Kolb's experiential learning theory (1984) serving as the framework around which this research was designed. In order to help our trainees construct their own knowledge and improve their technical skills through a transformative experience, this research employed an evidenced based approach to help guide resident participants through the four different types of abilities required to learn. Providing our participants with a meaningful concrete experience, guiding them through the process of reflective observation, helping them better understand the process through abstract conceptualization, and allowing them to engage

in meaningful opportunities for active experimentation was the ultimate goal of this lab. In order to meet this goal, however, attention to the following concepts was required.

Distributive practice and the deliberate practice effect played a significant role throughout this research project with both the setup and structure of this anastomosis lab. Distributive practice, as has been previously discussed, has been shown to result in improved acquisition and transfer of technical skills learned in simulation settings (Dawe et al., 2014; Moulton et al., 2006), and allows learned skills to consolidate through sleep between practice sessions (Louie & Wilson, 2001). Deliberate practice, as advocated by Ericsson (Ericsson, 2004), stipulates that practice sessions should be structured around well-defined learning objectives, include detailed feedback on performance, and be guided by error correction and opportunities to improve performance through repeated practice. These concepts were supported throughout this research by holding simulation lab sessions each week for the entirety of this simulation lab, with required attendance on the part of the junior surgical resident participants.

These weekly sessions were all held in three-hour blocks to accommodate participant schedules, with residents typically spending at least one hour in the lab during these sessions. This time allowed each participant repeated practice performing their own end-to-side anastomosis and also providing them with time to serve as an assistant. While optimum practice schedules to improve performance vary from one domain to the next, research supports practice sessions lasting an hour or less to be optimal for learning (Stefanidis & Heniford, 2009; Van Dongen et al., 2011). Consistent with this, Ericsson found that expert performers typically engage in practice without rest for an hour a day, especially in the early morning, and that concentration is the main factor in this time

constraint (Ericsson, 2006). While we did our best to hold morning practice sessions, we did have to schedule practice sessions in the afternoon in order to accommodate resident schedules.

Benchmark demonstration videos were provided to all participants, after the post-test, which provided them with on-demand access to compare their performance to that of the expert. This video was provided to serve as a motivating factor and should have also contributed to deliberate practice throughout the six weeks, helping participants set realistic performance goals for their next session. Using the analytics provided by the private, online, video streaming platform used for video distribution to participants, the benchmark video released at the end of week one was viewed a total of seventy-seven times over the course of the next six practice weeks. This averages out to approximately five views per participant. A second benchmark video was released at the end of week two (**view here**) to help clarify questions that most of the resident participants had around completing the transition stitch. This video accumulated a total of 40 views over the following five weeks, averaging out to an additional three views per participant. While the number of times participants viewed their own videos for playback analysis across the trial weeks ranged from five to twenty-eight views in total, for an average of twelve views per participant, this ultimately had no effect in post hoc analysis. When participants who viewed their videos ten times or less over the course of the seven weeks (n = 6) were compared to participants who viewed their videos more than ten times (n = 7), there were no statistical significance between these two groups in any of the seven variables recorded throughout this research.

The structure of this lab was also dictated by the tenants of motor learning theory, particularly with regards to the use of providing feedback to participants throughout their scheduled sessions. While Adams Closed Loop Theory of motor learning (1971) has been supplanted by Schmidt's Schema Theory (1975), his emphasis on the role of feedback, error detection, and error correction on psychomotor development remains poignant as it relates to this research. Adams stipulated that in order to learn a correct movement, "the subject needs knowledge of results to inform him about the correctness of the last movement, and response-produced feedback stimuli to inform about the progress of the current movement" (Adams, 1976, p.90). Building off Adams' theory, Schmidt later focused his efforts on determining the factors that most significantly influenced motor learning. In addition to concentration and motivation, Schmidt (2004) noted that an instructor's extrinsic feedback regarding errors "is one of the more important sources of information" (p. 305), especially as it pertains to knowledge of results. This finding has also been echoed in other research where immediate feedback has been found to be especially useful in correcting inappropriate actions related to procedural and motor skills (D. I. Anderson, Magill, & Sekiya, 2001; J. R. Anderson, Conrad, & Corbett, 1989; Grillo, 1999; Mory, 2004).

In order to implement this type of feedback in this vascular anastomosis simulation lab, the attending surgeon who served as a facilitator served a dual role in providing extrinsic, immediate feedback to the resident participants from both the Coaching and Non-Coaching groups, in an attempt to improve their technical skill. During the open practice sessions in weeks two through six, guided instruction and feedback was provided in general terms to all participants, in line with the specific

learning objectives of the anastomosis lab. Most of this directed feedback was provided

when specific questions arose on the part of the participants. For example, one resident,

still trying to master the transition stitch during week three, asked the facilitator to watch

them perform this part of the anastomosis and help walk them through it. While this

feedback was provided to the resident, the other resident participants in the lab also

benefitted from the discussion that followed.

In addition to this generalized instruction, the Coaching group had the additional

benefit of receiving specific, individualized, feedback from the facilitator based on a

video review of their previously recorded performance. During these coaching sessions,

the comments and discussion was structured around the Re-GROW coaching framework,

and guided by the following questions:

- **Review/Evaluate:** How do you think you did with the anastomosis during this

   past week? Are you satisfied with your performance?

- **Goal:** What do you want to work on this week with regards to your performance?

- **Reality:** What do you think you are doing well? How did you do on the scoring

   evaluation? Where were your deficiencies?

- **Options:** How would you like to improve on these deficiencies? If you could only

   work on one aspect of your performance during the next practice session, what

   would it be?

- **Wrap up/Way forward:** Are there any challenges to this plan moving forward?

   How can we best work through them?

Residents in this coaching group also had time to address any other questions they

may have had related to performing this anastomosis during these structured sessions. To

help reinforce this feedback, common themes that arose in these coaching sessions were combined in a summative email after each coaching session and sent to the participants (n = 6). The email sent after the first coaching session included the following reinforcing feedback:

1. Make intentional throws. Try not to drive the needle through the graft, only to realize it wasn't exactly where you wanted it, and then remove the needle to make another stick. This is going to increase your leak rate/cause more bleeding in the operating room.

2. Economy of motion. It's impossible to rotate the patient in the OR to make your suturing easier. Try to get comfortable rotating your body to make your throws easier. You are not confined to one position.

3. Make sure your assistant is providing enough tension on the graft. This allows you to better see what you are doing/improve placement of your bites. Also remember to ensure your assistant is following with the suture to prevent laxity in your closure.

For comparison, and to demonstrate the progression in type of feedback to the coaching participants, the summative email sent after the last coaching session (prior to the Post-Test) included the following reinforcing feedback:

1. Most of your times have improved substantially as your economy of motion has improved dramatically. Remember to fight the urge to grasp the needle with your Gerald forceps, use your Castro for this. As Dr Carruthers pointed out, use your Gerald as a platform to steady your Castro as you grab the needle.

2. You have all done a better job grasping the transition stitch around the heel of the anastomosis. Try not to travel too far as you complete this horizontal mattress suture, as this seems to be the area where most of your significant "leaks" are occurring.

3. You are all doing a much better job with intentionally placing your bites, compared to the first couple of weeks where there were a lot of sticks that ultimately were redone for better suture placement. Remember, your bites should be 1mm, no more than 2mm, apart with each throw. Making smaller, more frequent bites not only increases your time, but also produces more needle holes for leaks.

4. Remember to inspect your anastomosis regularly to make sure your sutures remain under tension and the graft lines up appropriately. It is tough once you get into a rhythm to stop and check things over but try and do this at least a few times throughout the anastomosis, you'll be pleased with your results.

5. Continue to use your assistant to your advantage. Don't let them be a passive "grasper." Tell them how you want them to hold/orient the graft. Force them to put appropriate tension on the graft and open up your graftotomy to allow for more precise suture placement.

This was the specific intervention strategy developed for this anastomosis lab using the best available evidence surrounding distributive practice schedules, the deliberate practice effect, reflective practice, and motor learning theory with respect to psychomotor development and technical skill acquisition. Further discussion of the

results will now be guided by the specific research questions that formed the basis of this study.

RQ1A: *Based on expert evaluator scoring using a validated technical skills checklist, to what extent do residents who participate in structured, individualized, video-based coaching attain technical skill, compared to residents in a self-assessment, video analysis only group?*

RQ2: *Based on expert evaluator scoring using the validated technical skills checklist, do residents who participate in structured, individualized, video-based coaching attain technical skill at a different rate than residents in the self-assessment video analysis only group?*

This research was unable to show a significant difference in technical skill acquisition between the structured, individualized, video-based coaching group, compared to residents in the self-assessment, video analysis only (Non-Coaching) group. Before this discussion continues, however, it is important to address the week one Pre-Test data analysis. Upon inspection, and despite participant randomization, Table 4 identified three variables that, when analyzed by an Independent Samples T test, differed significantly between the Coaching and Non-Coaching groups, prior to any intervention being carried out. The three variables in question included; Attending GRS (14.25 vs 17.29), Time to Complete the Anastomosis (24.17 vs 27.14) and Combined Technical Score (-34.67 vs 5.29). Given this information, the question becomes, how should we account for these differences between groups? Should we consider them significant at all? And do these results alter the way we interpret the final results from week seven (Post-Test)?

It is common practice, particularly in clinical trials, to present baseline characteristics of participants, and then run statistical analysis on these groups to show the groups are comparable. Despite this practice, Altman (1985) notes that using statistical testing on baseline comparability should be avoided when proper randomization has occurred as any baseline difference between the groups has a twenty percent chance of occurring simply due to chance. He further states that "performing a significance test to compare baseline variables is to assess the probability of something having occurred by chance when we know that it did occur by chance" (p. 126). Taking this into consideration with respect to this research, we should, however, question why these three variables may have produced a significant difference.

The first question that needs to be addressed, now that Table 4 has shown evidence of baseline differences, is did randomization occur correctly for this research? Elkins (2015) point out that when true randomization is not undertaken, practices such as quasi-random allocation, where participants are assigned based on age, birthdate, odd versus even number, etc., can contribute to systematic differences between groups at baseline. As a random number generator was used in this research to allocate participants to either Coaching or Non-Coaching groups, we should have avoided this potential confounder. The next, and perhaps the most important question is, whether or not the sample size (n =13) contributed to the significant baseline differences noted. To this question, the answer is most certainly yes. A small sample size, like this one, is much more prone to baseline imbalances due to random chance, compared to larger trials with more participants. This point is best made clear by Roberts & Torgerson (1999), when they point out "as the trial size increases, the absolute size of imbalance in baseline

characteristics will reduce, owing to a reduction in sampling error. Hence the absolute magnitude of any chance bias in outcome will tend to decrease with sample size" (p. 185). Because this research only employed thirteen participants, I believe this is what ultimately contributed to the differences of significance that were observed in the Pre-Test.

Finally, with regards to the baseline differences, it is hard to quantify exactly how much previous exposure each resident had to suturing small anastomoses like this one, which could have ultimately contributed to differences in time to perform the anastomosis and, more specifically, individual technique. While all participants were junior surgical residents in their first month of their second year of residency, previous suturing exposure and practice up to that point is extremely variable and could have also contributed to baseline differences. The four resident participants outside of the general surgical specialty (plastics and urologic surgery), although equally distributed across the Coaching and Non-Coaching groups, will also have had some differences in practical exposure to suturing. For this reason, I think these baseline differences are worth mentioning, and I am reporting statistical significance in Table 4 for completeness and transparency of the data. It should be noted, however, that the Consolidated Standards of Reporting Trials (CONSORT) (Moher et al., 2010) does not recommend baseline significance testing as common practice, as they regard it as "superfluous and can mislead investigators and their readers" (p.17). Instead, the recommended practice is for tables to be provided on baseline participant characteristics, allowing readers to use their own judgement to decide if any differences among participants are substantial enough to have potentially influenced the outcome of the research.

Despite these baseline difference between groups, the most important findings of this research relate to the improvement in technical skill experienced by both the Coaching and Non-Coaching group over the course of the seven-week vascular simulation lab. As measured by two expert evaluators using a validated scoring system, significant improvements were appreciated in both groups as evidenced by the Attending MOSAT and Attending GRS scores. With respect to research question two (RQ2), there were also no significant difference in the rate of technical skill acquisition between the two groups at weeks three, five or week seven (Post-Test). It appears that both the Coaching and Non-Coaching group made steady improvement in performance each week. This can best be appreciated in Graph 10 and Graph 11, where the rate of technical skill acquisition appears to run parallel between the two groups. This also held consistent in the within-subjects effect as the neither the interaction between attending MOSAT/attending GRS and Coaching/Non-Coaching reached statistical significance.

While there were no significant differences appreciated when comparing technical skill acquisition between the Coaching and Non-Coaching group, it was encouraging, despite our previous discussion on baseline differences observed in the Pre-Test, to see that there were no significant differences observed between groups in the week seven Post-Test (Table 28). This argues that any differences that may have actually been present at baseline favoring the non-coaching group were eradicated through our intervention, allowing the coaching group to "catch up" with the non-coaching group by week three and beyond.

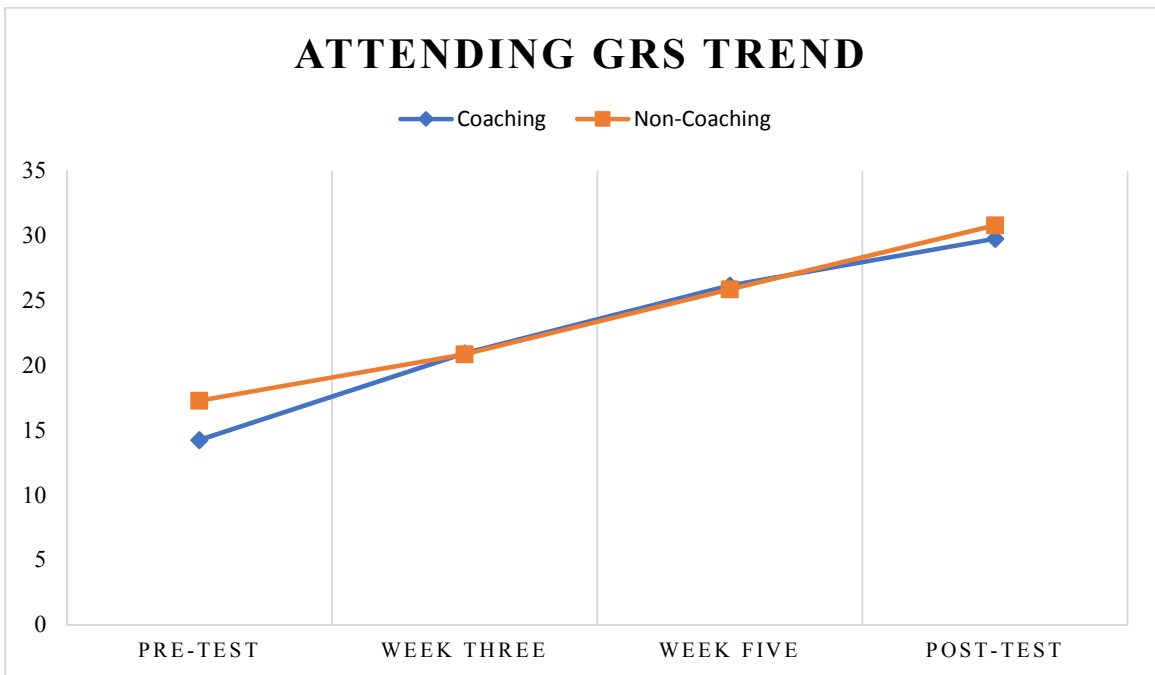**Graph 10:** *Attending MOSAT Group Trend*
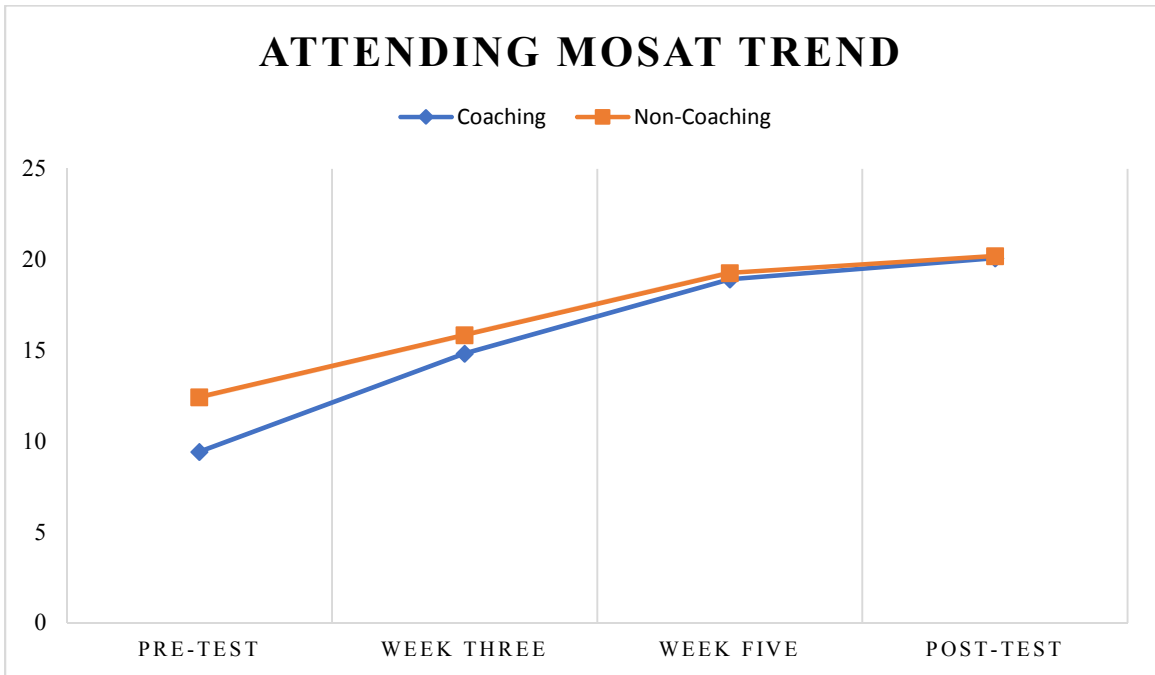
Graph 11: *Attending GRS Group Trend*

Table 28: *Week 7 Independent Samples Test*

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Attending MOSAT Week 7	Equal variances assumed	.082	.780	-.409	11	.690	-.1310	.3199	-.8351	.5731
Resident MOSAT Week 7	Equal variances assumed	13.288	.004	-1.495	11	.163	-2.357	1.577	-5.828	1.113
	Equal variances not assumed			-1.398	6.001	.212	-2.357	1.686	-6.482	1.768
Attending GRS Week 7	Equal variances assumed	.370	.555	-1.252	11	.236	-1.0357	.8270	-2.8558	.7844
Resident GRS Week 7	Equal variances assumed	.033	.860	-.421	11	.682	-1.667	3.960	-10.383	7.049
Averaged time Week 7	Equal variances assumed	.002	.962	2.185	11	.051	3.976	1.820	-.029	7.982
Leak Rate Week 7	Equal variances assumed	.685	.425	.569	11	.581	2.857	5.024	-8.200	13.914
CTS Week 7	Equal variances assumed	.018	.896	-1.500	11	.162	-9.429	6.285	-23.261	4.404

While the mean difference between the two groups was 3.01 for the MOSAT and 3.04 for the GRS in week one, the mean difference in week seven was 0.13 for the MOSAT and 1.04 for the GRS. It is also interesting to note that the MOSAT scores appeared to level out by week five for both groups, with significant pairwise comparisons evident between weeks one, three and five for both groups but not between weeks five

and seven. This was different than the GRS scores, which showed evidence of significant

pairwise comparison between all four test weeks. This effect may be explained by the

objectivity and binary scoring structure inherent in the MOSAT, compared to the five-

point Likert scale scoring of the GRS, which can lend itself to more subjectivity by the

evaluator. The MOSAT also had a top possible score of twenty-two points, while the

GRS had a top possible score of forty points. This larger variation in potential score for

the GRS may also have accounted for the significance appreciated across all weeks.

As the variables of time and leak rate are going to be discussed in greater detail

when we address research question four (RQ4), here they will be discussed regarding

their role in the calculation of the Combined Technical Score (CTS). The CTS was

developed for this research project as a way to objectively quantify resident participants

technical skill acquisition. While the CTS will need further research to validate this score

as a reliable measure of technical skill acquisition for trainees in a vascular anastomosis

lab, for the purposes of this research, this score represented the best way to combine the

measured variables into a score that would correlate with the Dreyfus Model of Skill

acquisition for each resident. In essence, the establishment of this CTS helped serve the

dual role of quantifying technical skill acquisition while serving as a criterion to help

identify their level of acquired skill. The CTS can be represented by the following

formula:

Combined Technical Score = 2 * (MOSAT * 2) + (GRS * 2) – Time – Leak

The ultimate goal for a simulation lab, such as this one, is to provide junior

surgical residents with the cognitive and psychomotor skills required to safely and

effectively perform procedures in a real-world setting. The Objective Structured

Assessment of Technical Skills (OSATS) was created in the late 1990's by surgical

educators and researchers at the University of Toronto in an effort to formally assess

technical skills in trainees (Hatala, Cook, Brydges, & Hawkins, 2015). By using direct

assessment, combined with a task-specific checklist (Martin et al., 1997), the goal was to

move away from the underlying assumption that individuals who have performed a set

number of procedures are technically "competent." Setting this fixed number of

procedures to dictate what the literature refers to as "competency" is not effective, as it

ignores the variability in individual psychomotor development. While proficiency-based

training, as opposed to time-based training, is now well established in the literature

(Ahlberg et al., 2007; Korndorffer Jr et al., 2005; Stefanidis et al., 2005) and has been

adopted by most residency programs, the definition of what constitutes competent skill in

a trainee has remained elusive for surgical training programs thus far. Up until now,

research evaluating the implementation of standardized approaches to help differentiate

between competent and noncompetent performers has been lacking (Szasz et al., 2015).

This particular research employed the Dreyfus Model of Skill Acquisition in an effort to

address this need and help identify the level of skill our resident participants were

ultimately able to achieve by participating in this structured vascular anastomosis lab.

Based on the level of experience these junior resident participants had with

performing an end-to-side vascular anastomosis prior to the start of this lab, all of them

would correlate with the novice stage of skill acquisition. This stage is governed by

learners who seek out strict rules to complete the task and have a limited ability to

prioritize and synthesize information (Carraccio et al., 2008). Dreyfus notes that these

learners do best when the task is deconstructed (S. E. Dreyfus, 2004) in order for the task

to make sense. While this research was strictly quantitative, the PI was able to note the

difficulty all participants had when trying to perform this anastomosis during the week

one Pre-Test. Most resident participants called on their assistant for help to try and work

through the task, as they had not been given any instruction or benchmark videos at this

point. Out of all thirteen resident Pre-Test videos, only one resident prepared the

anastomosed graft appropriately by spatulating the end prior to suturing it to the recipient

graft. A few residents performed the anastomosis with interrupted sutures, and two of the

anastomoses nearly fell apart upon completion, with leak rates of 100ml and 97ml

respectively. For this week one Pre-Test, mean CTS for the Coaching group was -34.67,

while the Non-Coaching group mean CTS was 5.29, with a combined group mean CTS

of -13.15. This, again, was the only week a significant difference existed between

residents in the structured, individualized, video-based coaching group, compared to

residents in the self-assessment, video analysis only group. No other weeks during the

anastomosis lab demonstrated a significant difference in CTS.

By week three of the lab, resident participants had exposure to the facilitator's

instruction, video benchmarks demonstrating a proper anastomosis and had completed

the Pre-Test evaluation as well as one practice week. The MOSAT checklist and video

benchmark was instrumental in deconstructing the task for residents, as was evident in

their performance videos for week three. This week would still be best represented as the

novice stage of skill acquisition, as all residents were still operating based on the strict

rules Dreyfus refers to and were very cautious in their anastomosis construction. This

was best evident by the increase in overall time taken to complete the anastomosis in

week three, as compared to the Pre-Test in week one. The Coaching group had a mean

time to complete that anastomosis of 30 minutes in week three, up from 24 minutes in week one, while the Non-Coaching group had a mean time to complete the anastomosis of 29 minutes in week three, up from 27 minutes in week one.

Resident performance during week three was also based on some basic knowledge of results, as was described by Adams theory of motor learning (Adams, 1971). The information resident participants had received in the first two weeks regarding their performance, along with the benchmark videos available to review and compare their performance with, likely helped them correct some of their basic technical errors, such as instrument handling and suture placement. Evidence of this can be seen in the slight increase of CTS this week, with the Coaching group improving from -34.67 in week one, to 5.00 in week three, and the Non-Coaching group seeing an improvement from 5.29 in week one, to 21.86 in week three. While there was no significant difference between groups, mean improvement of CTS from -13.15 to 14.08 was a significant difference from the week one Pre-Test.

Week five of the trial produced some of the best overall scores and was the week residents finally seemed to display the proper psychomotor development with regards to constructing this end-to-side anastomosis. Akin to the Dreyfus Advanced Beginner Stage of skill acquisition, resident participant's practical experience with this anastomosis and pattern recognition at this stage was definitely a factor in their overall progress. This research would agree with Benner (Benner, 2004) when she noted the "advanced beginner has a heightened awareness of any feedback on performance and pays close attention to the practice of colleagues" (p.193). This was evident in the resident participants actively seeking out feedback in performance. In fact, the request in week

two, on behalf of the residents, regarding proper placement of the transition stitch led to the production of our second benchmark video. The video performances for week five reinforce the fact that participants had reviewed this video multiple times, had practiced this transition stitch in the lab, and were now filtering information and feedback they had previously received, in an effort to produce a quality end-product. Week five was the only week the Coaching group, who had a mean CTS increase of 52, for a total CTS of 57, outscored the Non-Coaching group, who had mean CTS increase of 27.72 for a total CTS of 49.57, although this difference was not significant. Evident from reviewing week five videos, however, was a clear overall progression in both technical skill and psychomotor development.

Week seven (Post-Test) results are the closest we would be able to come in this lab setting to the level of skill Dreyfus labeled as competent performance. Efficiency was not preserved, as most resident participants increased their time in performing this anastomosis. What they traded in efficiency, however, they made up for in technical performance. Dreyfus and Dreyfus (1986) mention this as they describe the competent performer as becoming more vested in the resultant outcomes based on their actions. This could be seen during the week seven Post-Test as multiple participants verbalized wanting their anastomosis to be perfect for this final week. For example, when a stitch was throw that didn't fall as they initially had intended, they weighed the decision as it related to re-doing the stitch or making an alternative plan for the next throw to compensate on how the graft would ultimately lie as a completed product.

The week seven Post-Test saw a final CTS of 60 for the Coaching group and 69.43 for the Non-Coaching group. While these between group means were not

statistically significant, it does identify the significant improvement (p <0.001) in CTS

scores from a mean of -13.15 during the week one Pre-Test, to a statistically significant

mean increase of 65.08 in the week seven Post-Test. Gallagher et al (2012) have stated

that the goal of any simulation training is to ensure that trainees have practiced to the

point of automation of basic psychomotor tasks. This automaticity is similar to what

Richard Schmidt's Motor Schema Theory describes as the formation of a Generalized

Motor Program (Schmidt, 1975). Schmidt outlined the four steps of storing information

included in the production of a goal-oriented movement. These steps include storing the

initial conditions, such as proprioceptive information about the positions of one's limbs

and body in space; Storing response specifications for the motor program, which include

alterations in movement, such as changes in speed and force; Storing the sensory

consequences of the response produced, such as actual feedback stimuli received from

sensory organs during a task; Storing the outcome information, such as the success of the

response in relation to the original intended outcome. The difficulty, however, for

training junior surgical residents, is determining how long it takes for this motor program

to develop and contribute to automaticity in practice. Older literature seemed to offer

ranges to determine "competency" for procedures, such as 50 to as many as 300 cases for

flexible gastrointestinal endoscopy (Cass, 1999), and between 10 to 50 cases for a

laparoscopic cholecystectomy (Moore & Bennett, 1995).

We know, however, that objective assessments, such as that used in this research

are the best way to document and determine competent practice. As surgical training

programs need proper documentation of resident participation in simulation activities and

the ability to quantify performance to determine competent practice, this research

provides us with some useful data. With regards to both the Coaching and Non-Coaching groups participation in this vascular anastomosis lab, we can use the CTS to help identify resident performance according to the Dreyfus model of skill acquisition. While this scoring system will need to be validated with future research and resident performance in next year's vascular anastomosis lab, the following scoring model based on CTS will be used to identify level of technical skill acquisition:

1. 0 - 39 = Novice

2. 40-59 = Advanced Beginner

3. 60- 79 = Competent

4. 80 - 99 = Proficient

5. ≥ 100 = Expert

Using this scoring system, we have intentionally set a range on a scale of 0-100 for most performers, as we feel this best represents the stages that residents will move through during this simulation lab. We have also intentionally set up this scoring system so that an expert will score above 100 in the CTS. For example, the attending surgeon who facilitated this lab, and served as a coach, performed the anastomosis in 14 minutes and experienced a leak rate of 28ml. Assuming an expert would score a complete 22 on the MOSAT and 40 on the GRS, the CTS for this attending would be 126. The goal of this lab, as Gallagher (2012) noted, is to move residents towards attainment of automaticity in psychomotor performance. This should be accomplished in the Competent stage of skill acquisition. It should be noted that the automaticity we are speaking of is what Adams (1971) would describe as a closed loop model of motor learning. This automaticity exists in a simulated environment, under constant,

unchanging conditions. This level is different from that of a proficient or expert performer who, in a real-world, ever-changing environment, is able to maintain this automaticity at first by analytically thinking about their situation (proficient) and then, with more experience, will be able to perform this anastomosis in the most critical of circumstances without thinking about their actions (expert). This year's participants reached a mean CTS of 65.08 by week seven, which is just within the competent stage of skill acquisition, based on the scoring system we have proposed. Moving forward, our program can now use this CTS for future vascular anastomosis labs in order to serve two purposes. The first is to identify when a resident's performance is consistent with competent practice, which will allow them to participate in the operating room assisting an attending (expert) with this anastomosis. In the same respect, the CTS may also be used to help identify non-competent practice at the end of a scheduled seven-week lab. This designation would mandate that any resident who scored under this threshold would need to attend more practice sessions, until a CTS of 60 or better was attained, before they would be allowed to participate in the operative environment.

RQ1B: Using the validated technical skills checklist, to what extent do resident's self-assessment scoring of their own technical skill differ between those who participate in the structured, individualized, video-based coaching group, compared to those in the self-assessment, video analysis only group?

As previously discussed, Boud (2013) defined self-assessment as ''the involvement of students in identifying standards and/or criteria to apply to their work and making judgements about the extent to which they have met these criteria and standards'' (p. 12). This definition was instrumental in guiding the methodology of this research as it

pertained to evaluating the ability of junior surgical residents to self-assess their own technical skill acquisition. While the majority of evidence suggests that physicians have very limited ability to self-assess (Davis et al., 2006), multiple efforts have been made to try and improve this ability through the provision of benchmarks or explicit anchors for evaluation criteria (Ward et al., 2002). More recent research performed by Bonrath et al. (2015), demonstrated better self-assessment skills for residents who were assigned to a surgical coaching group. The research presented here, however, was unable to replicate this finding. As reported in chapter four, both the Coaching and Non-Coaching groups showed significant overall improvement in MOSAT and GRS self-assessment scoring across weeks one, three, five, and seven, but there was no between groups significance in self-assessment scores appreciated. When broken down by groups, the correlation coefficients with Attending MOSAT and GRS scores for the Coaching group ranged from -.580 to .429, while correlations for the Non-Coaching group ranged from -.604 to .687. This variability is likely due to the inherent flaws in methodology that arise when trying to evaluate self-assessment, particularly regarding the validity and reliability of the gold standard score, which will be discussed in greater detail with respect to research question three (RQ3).

Despite this weak to moderate correlation with Attending MOSAT and GRS scores, resident self-assessment scores for both the Coaching and the Non-Coaching groups significantly increased across weeks. Mean scores for the Resident MOSAT increased from 11.77 in week one to 19.77 in week seven, while the mean scores for the Resident GRS increased from 15.85 in week one to 28.23 in week seven. These results appeared to level out by week five (mean MOSAT 19.38, mean GRS 26.62), with no

significance difference noted between weeks five and seven. Comparative analysis using an independent samples T test also showed that self-assessment scores between the Coaching Group and the Non-Coaching group were not statistically significant from one another across any week. One of the reasons Bonrath et al (2015) may have been able to show a difference in self-assessment scores for residents who were surgically coached is because only the residents randomized to the coaching group had the ability to view video playback analysis of their performance. This is an important point, as this video benchmark of their performance likely contributed to reflective practice, as the authors point out, which may have allowed them to better assess their own performance. The research presented here, however, provided *both* the Coaching, and the Non-Coaching groups with an opportunity to not only evaluate their own performance using video analysis, but also compare their performance to that of an expert. The steady improvement in Resident MOSAT and GRS self-assessment scores observed in this research may be best explained through what Donald Schön refers to as both reflection-in-action and reflection-on-action (Schön, 2017).

Reflection-in-action is a cognitive real-time approach one uses to analyze a situation and adapt our thoughts and actions to the requirements of the change one is trying to achieve. When engaging in reflection-in-action, one will stop during the action, make adjustments, and alter methods to improve their practice as necessary. Schön (1987) further clarified this: "What distinguished reflection-in-action from other kinds of reflection is its immediate significance for action" (p.29). This specific type of reflection was frequently witnessed over the course of this research, particularly during the practice sessions. As resident participants were trying to complete their anastomosis, discussion

would often take place between residents over the correct steps of the procedure, especially during the first few weeks of the anastomosis lab. Suture placement was discussed extensively during the early weeks, especially with regards to weighing the effects of a wrong throw of the suture against redoing the suture throw altogether in an effort to correct suture placement but potentially impacting the leak test by making an additional hole in the graft.

Reflection-on-action, however, is a post-analysis of our action once we have completed the activity (Eva & Regehr, 2005). Its purpose is to reconstruct the experience, based on what we can remember, and organize the experience with the purpose of understanding what happened during it, and what lessons we can ultimately draw from participating in the experience. The goal of this research was to provide resident participants with benchmark videos of an expert performing the anastomosis, against which they could compare their own video to, and reflect upon, in an effort to formulate goals and objectives to help improve their performance during the next lab session. Not only has this type of practice been shown to be effective, but literature has also shown that making mistakes and reflecting upon them in this type of safe setting is one of the best ways to learn (Agha & Fowler, 2015).

Benchmark videos alone, however, may not have been the only factor contributing to the reflective practice in our trainees, as a coach or a facilitator can also be instrumental in this process. This was the purpose of structuring this research around the POMM (Peno & Silva Mangiante, 2012) to help promote reflection, particularly in our structured, individualized, video-based coaching group. While we were unable to show a difference in self-assessment scores between the two groups, both showed

significant overall improvement in self-assessment scoring, and both were exposed to the coach/facilitator to varying degrees. For the Coaching group, reflection was specifically addressed in separate, open discussion that was structured around the Re-GROW coaching framework. In the weekly lab sessions, the facilitator openly discussed options to improve technical performance with all participants, which included discussions on comparing their performance with the benchmark video.

Although it was not measured, and is tough to control for, it can be speculated that both processes allowed for the coach/facilitator to play an active role in all of the participant's technical skill acquisition through both direct and indirect promotion of a reflective process. As Schön himself notes, "Every attempt to produce an instruction is an experiment that tests both the coach's reflection on his own knowing-in-action and his understanding of the student's difficulty" (1983, p. 104). While we know that both instructional methods (coaching and providing benchmark videos) have individually been shown to confer a benefit with self-assessment scoring in previous research (Bonrath et al., 2015; Martin et al., 1998), it becomes difficult to differentiate how much each methodology did or did not contribute to the improvement in self-assessment scores seen in this research. Future research may be directed at trying to measure the overall effect on reflective practice that can be established using coaching frameworks versus benchmark videos alone, or in combination, to promote self-assessment.

RQ3: Based on the validated technical skills checklist, to what extent do resident self-assessment scores of their technical skill correlate with expert evaluator scores?

As has been highlighted in this research, the popularization of simulation labs and other technologies, as a major component in technical skill acquisition and training, is a

direct result of the policy changes implemented in graduate medical education and

surgical residency. Simulation training comes at a cost, however, not only in equipment

and materials, but also in the time commitment required by faculty to staff these

simulation labs as facilitators and evaluators. This time commitment makes the process

significantly more resource intensive in an effort to benefit a select few residents

(Kneebone et al., 2010). In the vascular anastomosis lab used in this research, the

availability of three vascular surgeons was required throughout the seven weeks, one to

serve as the facilitator and coach to the participating residents, and two to serve as the

blinded video reviewers and evaluators.

The time demands to run this lab were substantial, with the facilitator spending

four hours each week for a total of five weeks overseeing the lab, in addition to reviewing

the coaching group videos. The video evaluations alone were a considerable time

commitment for the two blinded faculty, especially given the fact that each of the thirteen

resident videos ran anywhere from twenty to forty minutes long, with a total of fifty-six

videos produced over the course of seven weeks. Herein lies the ulterior benefit of

simulation-based training. Faculty demands could ultimately be lessened, to some extent,

if simulation-based activities did not have to be solely faculty driven, but could also be

self-directed by the resident participants themselves (Arora et al., 2011). Self-

assessment, in this respect, has the potential to serve the dual role of assessing one's own

learning and performance, while also helping to offset some of the time commitments

required by faculty.

This research, however, aligns nicely with what previous research in this area

suggests; simple correlation would allow you to believe that there is typically very little

agreement between self-assessment and expert assessments (Davis et al., 2006; Falchikov & Boud, 1989). Because of the consistency of these findings, it has become very difficult to know exactly how to implement self-assessment as an effective teaching modality in surgical training programs. Looking at the correlation coefficients obtained in this research (Table 29) comparing Attending MOSAT and GRS scores to Resident MOSAT and GRS scores over the four trial weeks, the following correlational trend was observed:

Table 29: *MOSAT and GRS Correlation Coefficients Weeks 1, 3, 5, and 7*

MOSAT Correlation Coefficients

Week 1	Week 3	Week 5	Week 7
$r_s = .180$	$r_s = -.165$	$r_s = .478$	$r_s = -.369$

GRS Correlation Coefficients

Week 1	Week 3	Week 5	Week 7
$r_s = .402$	$r_s = -.235$	$r_s = .448$	$r_s = -.389$

According to these results, the only weeks that reached a moderate correlation, based on a Spearman rank correlation, was the GRS evaluation in week one with $r_s = .402$, and MOSAT and GRS correlation in week five with $r_s = .478$ and $r_s = .448$ respectively. Outside of these weeks, the overall weak correlations observed are in agreement with the majority of the literature done on self-assessment. It is also worth noting that none of these correlations between Attending and Resident MOSAT and GRS scores achieved statistical significance. What this means for this particular research is that we cannot be sure (>5% chance) that the strength of the relationship did not happen due to chance alone, and perhaps this is the reason we had a moderate correlation in week one GRS ($r_s = .402$), while the week one MOSAT was very weak ($r_s = .180$). This, again, is a problem encountered due to the small number of participants recruited for this study, as statistical significance is more dependent on the sample size than the

correlational coefficient. Despite this, however, there are some positive attributes within this particular data set. The moderate correlations appreciated in week five (r_s = .478, r_s = .448), although not statistically significant, could be indicative of a trend towards the residents becoming better at their own self-assessment over time.

Contrary to the research performed by Rizan, Ansell, Tilston, Warren, & Torkington (2015), providing participants with benchmark videos for this anastomosis lab did not appear to improve the overall correlations between resident self-assessment and faculty expert assessment. I do not think, however, that our data shows that resident self-assessment is lacking or was not a factor in resident's overall improvement in technical skill acquisition across the seven weeks of the lab. This researcher would agree with Ward et al., (2002) that there exists the potential of a flawed methodology when assessing this relationship through simple correlational analysis. One of the problems the authors identify as a confounder is related to the assumption that the reliability and validity of using experts to serve as the gold standard should be viewed without skepticism. Here they note "even if one assumes that clinical supervisors are, in fact, accurate (valid) judges of cognitive achievement, any conclusion with respect to the accuracy of self-assessment presumes that experts are providing a fair measure of clinical performance" (p.68). If we look at the initial data from this research, the Kappa coefficient between the two expert evaluators was, in fact, very weak at 0.12 for MOSAT and 0.01 for GRS. The concordance between these scores, however, was high for Pearson Correlation (0.86, 0.80), Kendall's Tau (0.70, 0.66) and Spearman rho (0.84, 0.83). This occurred because although one of my evaluators was stricter at grading than the other evaluator, their scores both correlated with one another. In other words, as one

evaluators score increased, so did the other, and vice versa, even if the actual number was different. Because of this concordance, it was decided that the average of the two scores be used as the "gold standard" score for each resident. Using a third evaluator could have made this evaluation process more robust, and perhaps may have improved my inter-rater agreement, but this was not possible due to faculty availability.

Another factor Ward et al. (2002) identify as a methodological issue that can arise when comparing self-assessment scores to those of an expert is the assumption that all students are evaluating themselves by tapping into the same aspect of their performance. These *within group* differences have rarely been evaluated in research but should be accounted for. We did not formally review each question on the MOSAT and GRS evaluation form with the resident participants prior to them filling out these evaluations. This could have helped clarify the particular aspects of technical performance we were evaluating, specifically on the GRS, as this evaluation left more room for performance interpretation than the MOSAT.

Despite this overall weak correlation between faculty and residents, there was a consistent trend in the improvement of self-evaluation scores over time, particularly by week five (MOSAT = .478, GRS = .448). This unfortunately changed in week seven, with the scores producing a negative correlation, which means that the variables are inversely related, signifying that as one score increased, it's correlated score decreased. This negative correlation could be due to the observation that residents were, overall, much less satisfied with their results for the Post-Test compared to their week five performance. Going forward with this lab, it would be a suggestion to not label the final week as a Post-Test, and simply refer to it as week seven. Labeling it as a Post-Test

seemed to add a degree of pressure on the participants that was not present in the preceding weeks. The participants described being more nervous, as if they were actually taking a test during this week, more so than the prior weeks. This may have created more anxiety on the part of the participants, causing them to feel like they were making more overall mistakes, in addition to adding to their anastomosis time and increasing their leak rates, which was ultimately reflected in the way they self-evaluated themselves.

Given the significant improvement displayed by residents from Pre-Test to Post-Test and the overall positive trend in resident self-evaluation scores that appear to trend somewhat with faculty evaluation scores, findings from this research would include the recommendation that this aspect of self-evaluative performance review be maintained in future vascular anastomosis labs. A similar research project evaluating the role of self-assessment in a vascular anastomosis lab also found self-assessment and expert assessment correlated poorly, yet the authors concluded that "self-assessment with expert feedback throughout training appears to offer an efficient method of improving the technical performance of surgical trainees as an integral part of a structured surgical training program" (Pandey et al., 2008, p. 289). The research presented here would agree with this statement and our institution will continue to employ this combined modality in an effort to improve technical skill acquisition in junior surgical residents. A future research project would be to explore the role of peer evaluations on technical performance, which has previously been shown in several disciplines, including surgery, to be more accurate than self-assessment (Falchikov & Goldfinch, 2000; Risucci, Tortolani, & Ward, 1989). This could be another valuable strategy in helping to gauge

performance improvement in junior surgical residents and a potential way to help reduce the faculty resource utilization these simulation labs require.

RQ4: To what extent do the current posttest scores on time to fashion anastomosis and leak rate compare with the prior three years of resident vascular anastomosis scores?

The goal of this fourth research question was twofold: First, because of the variables collected as part of this research (time and leak rate), it provided us with a potential way to compare resident performance from previous vascular anastomosis labs at our institution with this year's lab, which employed a new format including benchmark videos, video playback analysis, and a validated scoring system, both with and without coaching. Second, depending on the analysis, it could provide us with some insight into the outcomes our department had been measuring previously, to see if we should permanently alter the format of our annual vascular anastomosis lab moving forward. This research question ultimately proved to be one of the more interesting findings of this research after final analysis, as will be discussed.

Our department of surgery previously published results in 2012 attempting to identify objective procedural end-product metrics for surgical residents participating in a vascular anastomosis simulation lab (Okhah et al., 2013). The two main metrics that were identified in this research were time to complete the anastomosis and saline leak rate. In addition to these performance measures, four technical errors including suture technique, locking sutures, air knots, and broken sutures were also identified as a way to track performance during the course of this vascular anastomosis lab, which consisted of three practice weeks, in addition to the Pre-Test and Post-test. Based on this published data, in an attempt to quantify any improvement in a resident's performance during their

participation in our vascular anastomosis lab, a simplified scoring system was developed

and implemented in 2013. While this score was developed with the intention to rank

residents according to performance, it ultimately became tough to use this score as a true

marker of technical skill. The first problem was the inconsistently and subjectivity in

deducting points for each technical error made from the final score during the Pre and

Post-Test. For example, if a resident scored a 50 based on the equation "*Score = 300 –*

(Time x 10) – Leak Rate," and they also made three technical errors while performing

their anastomosis, their score would actually be 47. These technical errors were also

randomly chosen to be included in the analysis and had not been previously validated as a

marker of technical skill related to a vascular anastomosis. In addition to this, this

simplified score placed the majority of weight on the variables *time* and *leak*, which, as

we will discuss, may not be the best measures of technical performance.

     The second problem with this simplified score was the requirement that residents

be limited to only thirty minutes to complete their anastomosis. As there are a significant

percentage of residents who ultimately take more than thirty minutes to perform this

anastomosis (30% of residents in the 2019 class), this time constraint always prevented a

percentage of residents from completing their anastomosis, or caused them to rush and

make more technical errors, which would then alter their final score. The third, and

perhaps most notable problem with this simplified score, was related to the data analysis

over the course of the last four vascular anastomosis labs (which included the current

2019 data). Although our current anastomosis lab participants showed significant

improvement in technical skill over the course of our anastomosis lab, based on attending

and resident MOSAT and GRS, if we based their score solely on the simplified score, the

opposite was true. 2019 resident participants appear to have regressed in their technical

skills over the course of this lab according to the simplified score, with a median decrease

of 33 in their simplified score from pre-intervention (*Mdn* = 117) to post-intervention

(*Mdn* = 84). The same was true when we compared the prior three years of simplified

score data (2016, 2017, and 2018) from the anastomosis lab using the Wilcoxon signed

rank test design. These results show that the only class to show a significant

improvement in technical skill was the 2017 class. This class demonstrated a statistically

significant median increase (*Mdn* 46) in resident's simplified score from Pre-Test (*Mdn* =

60) to Post-Test (*Mdn* = 106). It would appear, based solely on this simplified score

analysis, that residents who participated in our simulation lab over the last four years

improved in their ability to perform an end-to-side vascular anastomosis only 25% of the

time.

We now know, based on the objective data presented in this research, which was

structured around a validated technical skills checklist in addition to variables of time and

leak rate, that our residents do, in fact, improve their technical skill acquisition over the

course of a seven week vascular anastomosis simulation lab. The research presented

here, however, now calls into question the content validity of the simplified score.

Content validity relates to this simplified score's ability to accurately represent a

resident's technical skill acquisition, related to creating a vascular anastomosis, based

primarily on the factors of time to complete the anastomosis and measured leak rate.

While time is an important factor in any surgical procedure, this specific variable has

been looked at previously in the literature with regards to outcome. Datta et al., (2002)

attempted to objectively measure whether differences in a surgeon's manual dexterity had

any impact on a simulated vascular surgical procedure. While they found that trainees with better manual dexterity produced better outcomes with regards to anastomotic leak and degree of stricture, they also found that time taken to perform the procedure did not have any influence on these outcomes. In a similar manner, when Jensen et al., (2008) evaluated the effectiveness of a vascular simulation lab on resident's technical skill, they noted that both junior and senior level residents ultimately took more time to complete the anastomosis during the Post-Test. While the authors did not provide commentary on this result, having identified the same finding in our current research, we speculate that residents are intentionally paying closer attention to detail in fashioning their anastomosis during this Post-Test. This potentially caused them to slow their overall pace while suturing and contributed to the increased time that both Jensen's research and our own research documented.

Collectively, this data provides enough evidence to show that time to complete an anastomosis should not be used as a solitary measure of technical skill acquisition when assessing resident performance on a vascular anastomosis. In addition to this, the previous requirement of our vascular anastomosis lab stipulated that the anastomosis had to be created in thirty minutes or less. In our current research, the average time for both groups of residents to complete this anastomosis during the Post-Test was 27.69 minutes. Schwartz et al., (2014) performed similar research on junior residents in a vascular anastomosis lab and reported an average time of 35.6 minutes to complete the anastomosis after instruction. Their recommendation was to allow residents at least one hour to complete this anastomosis for the Pre-Test, then limit them to forty-five minutes for subsequent sessions. Our research would agree with this recommendation, as forty-

two minutes was the longest time it ever took one of our residents to complete this anastomosis, and this occurred during week five. While time to complete the anastomosis is important, and should be documented, it should only play a minor role in an objective assessment of a resident's technical skill, as this current research, and that of others, has shown.

The second major component of this simplified score, which had been used as a marker of resident's technical skill acquisition, is the anastomotic leak rate. While various models have been described to train residents in performing a vascular anastomosis, materials used typically include both real tissues such as porcine aortas, and artificial tissue like silicone, Gore-Tex, and PTFE (Jensen et al., 2008; S. Schwartz et al., 2014; Sidhu, Park, Brydges, MacRae, & Dubrowski, 2007; Wilasrusmee, Lertsithichai, et al., 2007). The most difficult aspect of working with artificial tissue, particularly the PTFE graft used in this research, is dealing with needle hole leaks. Needle hole leaks are a known problem that commonly exist in situations where prosthetic grafts are used (Baker, 1987). To overcome this limitation, companies have made sutures, such as polypropylene, with smaller needle-to-suture diameter ratios that approach 1:1, which were the sutures used in this research. While this ratio theoretically allows the suture to fill the needle hole, research has shown that although leak rates are reduced, they are still present (Dang et al., 1990). This can be a significant problem in real world settings, where vessel trauma and bleeding can compromise outcomes. To combat this, there has been significant research on how to best deal with the needle hole leaks in vivo, such as employing the use of surgical sealants such as fibrin glue, thrombin, and protein-based adhesives (Rogers, Turley, Cross, & McMonagle, 2016).

For purposes of this vascular anastomosis lab, we have to acknowledge that despite a near-perfect anastomosis, needle holes ultimately contributed to some amount of leak upon completion, despite proper technique. For example, when our attending vascular surgeon and lab facilitator performed his own anastomosis using the same materials as resident participants, the amount of saline leak recorded was 28ml. These needle-hole leaks can best be appreciated by viewing this example video where the straight lines of saline exiting the graft signify needle hole leaks (**play here**). While some researchers have tried to simulate the viscosity of blood during an anastomotic leak test, such as the one we used, by using vegetable oil (Datta et al., 2002), saline is easier to work with and more readily available in simulation labs. Saline will, however, flow more freely through these needle holes than would vegetable oil or blood. As measured with a wide-open incomplete anastomosis, the saline leak test used in our research produced a total of 100ml of saline over thirty seconds. For comparison, our resident participants experienced average leak rates of 63ml during the week one Pre-Test, 59ml during week three, 47ml during week five, and 48ml during the week seven Post-Test.

Given this information on time to complete the anastomosis and leak rates, this research has shown that a simplified score is not an accurate way to quantify a resident's technical skill acquisition, specifically as it relates to performing a vascular end-to-side anastomosis. Looking at the past years of resident data that was recorded in our vascular anastomosis lab, there is very little that we can take away from this. Our current resident participants displayed significant improvement in their technical skills as measured by an objective, validated scoring system, yet showed no improvement when evaluated by the simplified score. It is very likely that previous resident participants, particularly those

from 2016, and 2018, who showed no improvement in their simplified score from Pre-Test to Post-Test, also improved their vascular anastomosis technical skill over the course of the lab, but valid measures were not in place to accurately quantify this improvement. Thus, a simplified score cannot be used for measuring or quantifying a resident's technical skill in a vascular anastomosis lab, as it lacks the content validity required for this assessment.

The importance of this research question lies in the fact that we now have the local evidence needed to institute the more robust comprehensive technical score (CTS) for use in our vascular anastomosis simulation lab moving forward. The CTS has been shown, in this research, to be a more reliable indicator of our ability to measure and quantify the degree of a resident's technical skill acquisition across weeks. Its implementation in future vascular anastomosis labs may help us better track resident progress and degree of competency upon completing the lab, while also helping us identify residents who may need further practice in the simulation setting prior to assisting with this anastomosis in the operative environment. Further research is needed to validate this scoring system during the next vascular anastomosis lab in 2020.

Final Analysis

As this research was unable to show a significant difference between the two modalities implemented, the first major question we have to address regarding this research is what exactly contributed to the improved performance in technical skill acquisition experienced by residents in both the structured, individualized, video-based coaching group and the self-assessment video analysis only group? Having explored

each research question in detail and reviewing the descriptive and inferential statistics obtained from the lab, there is unlikely a single answer to this question.

We have to acknowledge the fact that there was going to be some improvement in performance simply based on resident practice alone. Even without proper instruction, residents would likely have improved their technique across the seven weeks to varying degrees. We also know that residents who participate in basic, standardized vascular simulation labs show improvement in overall performance, while those who adhere to a distributive practice schedule, such as that used in this research, can further enhance this performance (Moulton et al., 2006). These distributive practice sessions alone, however, cannot completely account for the degree of improvement observed in both attending and resident self-assessed MOSAT and GRS scores.

The benchmark videos provided to all participants, along with the playback analysis and video self-evaluation scoring also likely played a significant role in the technical skill acquisition that was appreciated by all junior surgical residents participating in this research. Based on the number of views both the benchmark videos and individual videos accumulated over the course of the lab, it would be hard to excuse this effect. While participant motivation and reflective practice was not measured in this research, the overall participant response to the structure of this lab incorporating these on-demand videos was overwhelmingly positive. How much these videos further promoted individual reflective practice, which has been shown to have a significant effect on technical skill acquisition (Stefanidis, Korndorffer, Heniford, & Scott, 2007), is not known. This is especially pertinent to the self-assessment, video assessment only group,

who showed similar improvement in performance as the structured, individualized, video-based coaching group.

The second major question we have to address is why this research was unable to show a statistically significant difference in technical skill acquisition between residents in a structured, individualized, video-based coaching group, compared to residents in a self-assessment, video analysis only group. As previously discussed, the use of coaching frameworks both with and without video analysis, as well as video-based self-assessment have been previously shown to improve technical skills in surgical settings to a greater extent than when these modalities are individually compared to standardized instruction (Bonrath et al., 2015; Farquharson et al., 2013; Singh et al., 2015; Soucisse et al., 2017). The one area in the literature that had yet to be explored at the time of this research project was comparing video-based coaching to self-assessment video analysis, to determine if one was superior over the other. As both modalities are unique and can each contribute to technical skill acquisition in various ways, especially with regards to reflective practice, we were likely unable to show a difference between the two modalities due to the low number of resident participants (n=13) enrolled in this research. As the control group only missed out on the additional video-based coaching and email reinforcement, this research was underpowered to detect a potential difference between the experimental and control group. Unless a large, multicentered trial can be coordinated, it is unlikely that a difference between video-based coaching and self-assessment video analysis will be able to be elucidated.

Another issue that may have contributed to this research's inability to show a difference between structured, individualized, video-based coaching, compared to

residents in a self-assessment, video analysis only group is resident interaction. While the researcher did their best to randomize residents and maintain segregation of the video-based coaching sessions outside of normal practice sessions, the researcher could not isolate both groups during practice and video sessions due to resident schedules and availability. Because of this, residents from the Coaching group may have acted as an assistant to residents in the Non-Coaching group and vice versa. While residents in the coaching group were encouraged not to talk about the specifics of the framework or coaching discussions with the control group, discussion held between residents during practice sessions (or outside of the lab) could have touched upon some of this methodology, or led to reflective discussions that may have eradicated any potential benefits conferred by video-based coaching.

Implications

Although this research was unable to show a significant difference between structured, individualized video-based coaching, and self-assessment video analysis only, there are some meaningful outcomes with regards to this research that can be inferred both locally and nationally. First, this research adds to the literature on the use of video analysis and coaching in surgical simulation labs, as a means to improve junior resident's technical skill acquisition. This research also adds to the literature on the ability of residents to self-assess their own performance in simulation labs. Although correlations between attending and resident scores were weak to moderate at best, resident self-assessment scores improved significantly across weeks. This research finding opens the door for more research into this area, particularly with regards to peer to peer assessment and evaluation in surgical simulation labs.

The local effects of this research are perhaps even more significant. As our fundamentals in laparoscopy (FLS), endoscopy (FES) and robotic curriculums are constantly being updated and refreshed with content, the benchmark, on-demand videos that were used in this research will now play a significant role in future simulation labs and curriculum development within our department. Residents who participated in this research have been a driving force in this movement, as they felt the videos helped them significantly throughout the vascular anastomosis lab. As a result of this, benchmark instructional videos on laparoscopic cholecystectomy and laparoscopic appendectomy are currently being prepared by the author of this research. Updated vascular anastomosis videos will also be created for next year's simulation lab. The vascular attendings who were not a part of this research, but who viewed the benchmark videos, now want to produce their own videos to help educate residents on how to perform two other vascular anastomoses, specifically an end-to-end anastomosis and how to properly parachute an anastomosis. The overarching goal, stemming from this research, is to continue building a local library of on-demand videos that residents may access at their leisure, that have been created and approved by local surgical experts, who can teach the key steps and proper technique of a specific procedure prior to a simulation or real-world experience.

Due to the overwhelmingly positive response by the residents who participated in this study, a recommendation for future research would be to include qualitative analysis into the study design. As mentioned previously, multiple residents from both the Coaching and Non-Coaching group commented throughout the lab as to how much they appreciated the methodologies employed and how much these videos added to their

simulation experience and technical skill acquisition. Having formal documentation of these resident responses would have added to the overall strength of this study.

This research has also contributed to the development of a comprehensive technical score (CTS), to be used in our vascular anastomosis lab, in an effort to quantify a resident's technical skill acquisition according to the Dreyfus model. While further research is needed to validate this score, it can now be used as a local guide to assess what stage of skill acquisition residents are at as they progress through the next vascular simulation lab. If, by the final week of the simulation lab, a resident score is below the threshold for competent practice (CTS < 60), they would be required to complete more simulation sessions until they are able to reliably achieve a baseline competency score. While this research was unable to show a difference between residents exposed to a video-based coaching framework compared to those in a self-assessment, video analysis only group, the fundamentals of the two modalities have proven to play an important role in surgical simulation and surgical training. This research has shown that both modalities can be easily incorporated into a simulation lab format and produce meaningful results. It will be the authors recommendation that our department provide residents with exposure to the Re-Grow coaching framework, combined with a self-assessment of technical skill, to help reinforce reflective practice in next year's vascular anastomosis lab.

Lastly, our previous anastomosis labs held within our department ran over an average of five weeks. After taking the pre-test and the post-test weeks into account, this allowed participants to have only three practice sessions with the facilitator to help improve their technique. Based on the resident performance data documented in this

research, where resident's technical skills did not start to plateau until week five, three practice weeks is unlikely to provide residents with enough distributive practice sessions to increase their performance to the level of competency designated in this research. While there is no agreed upon length of a simulation lab, as number or practice sessions do not confer competency, this author would advocate for no less than four practice sessions in between the Pre-Test and Post-Test, for a minimum total of six weeks in future versions of this vascular anastomosis lab.

Limitations

Despite the attempt at a rigorous research design, there are some limitations associated with this project. The first limitation is the single-centered research design employed. This single center design limits our ability to generalize the findings of this research to other surgical residency programs. The second limitation is the small number of residents enrolled in this trial (n=13). Because of this low number of participants, concern can be raised whether or not this study was adequately powered to detect a difference between the two independent variables. There was no way to increase this number, however, as this research was limited by the availability of second year residents in our training program who are required to attend this anastomosis lab. Multiple measures on each participant (four video session per resident) were collected as an attempt to help offset this limitation.

In addition to this small number of participants, and their friendly interaction outside of this anastomosis lab, the control group may become more highly motivated to perform at a higher level in order to be competitive with the experimental group. This phenomenon is known as compensatory rivalry and is difficult to control. There were

two residents who were assigned to the self-assessment video analysis only group who questioned whether or not they could be reassigned to the video-based coaching group during week one. After reassurance that they would receive all coaching-based materials and an opportunity to have their own one-on-one sessions with the coach and facilitator at the conclusion of research, they did express their satisfaction with this plan. In order to adjust for this compensatory rivalry however, a recommendation for future research would be to measure motivation in both groups and statistically adjust for this using an analysis of covariance (ANCOVA). Lastly, maturation is always a threat to internal validity in experimental designs. Over the course of the seven weeks during which this research takes place, participants should normally improve upon their technical skills over time simply due to practice. Random assignment of participants will help to offset this.

Conclusion

Surgical training has changed dramatically over the past two decades in light of concerns over patient safety, resident oversight, and resident well-being. Because of this change, residency programs have recognized that the cognitive apprenticeship model, pioneered by William Stewart Halsted in the early 20th century, and based on the "see one, do one, teach one" method, may no longer be the most effective strategy to prepare surgical residents for practice (Gallagher et al., 2012). While policy implementations by the ACGME, such as a reduced hourly workweek and competency based medical education, have been effective in addressing oversight and improving resident well-being, the cumulative effect of these implementations have called into question the technical competency of residents graduating from surgical training programs. As studies have

since shown a positive correlation between simulation training and technical competency in the operative setting, surgical training programs have placed a greater emphasis on using simulation bench model workshops to improve junior resident's performance on competency-based procedural skills, such as performing a vascular anastomosis.

This research attempted to evaluate if structured, individualized, video-based coaching improved technical skill acquisition to a greater extent than self-assessment, video analysis only in junior surgical residents attending a vascular anastomosis simulation lab. While I was unable to show a difference in technical skill acquisition between these groups, significant improvements in performance was witnessed over the course of seven weeks. Benchmark videos did not seem to improve the ability of resident's self-assessment of technical skills, as correlation with expert evaluator scores were weak to moderate throughout. This research was able to develop a scoring system to better quantify and define technical skill acquisition according to the Dreyfus model. While this score needs further validation, it may better help surgical training programs to define competency in resident participants prior to participation in clinical practice settings. This research adds to the overall literature on the development of psychomotor skill, video playback analysis, self-assessment, coaching, and competency-based performance as defined by the Dreyfus model of skill acquisition.
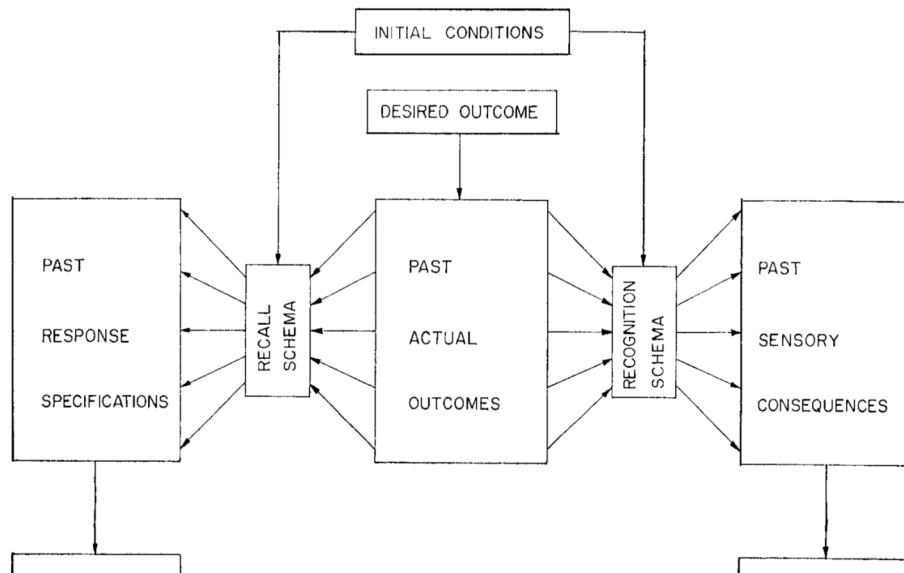
Appendices

Appendix A: Adams Closed Loop Theory of Motor Learning
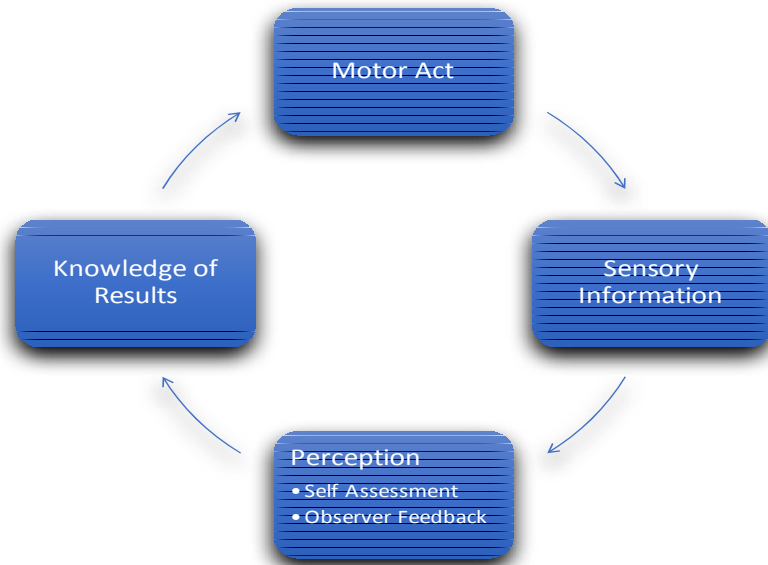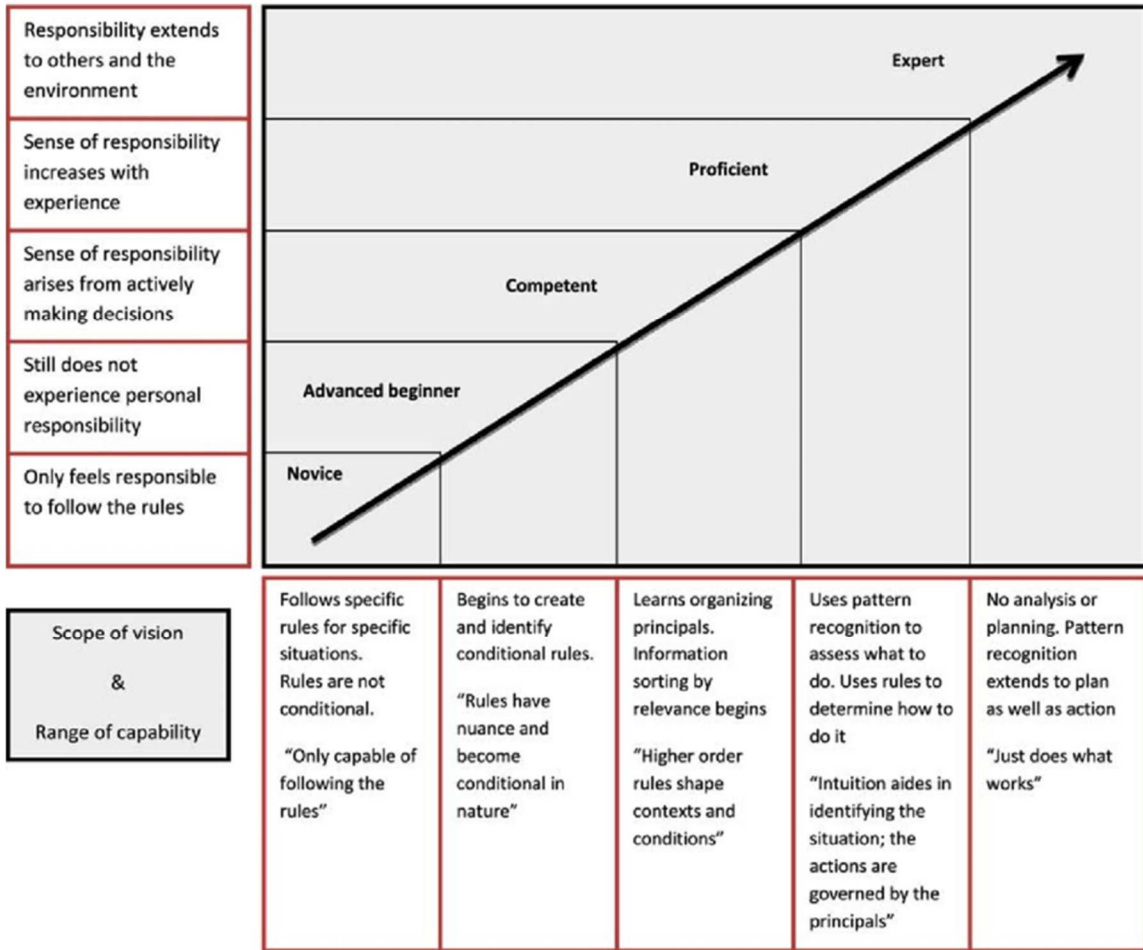

Appendix B: Schmidt's Recall and Recognition Schema
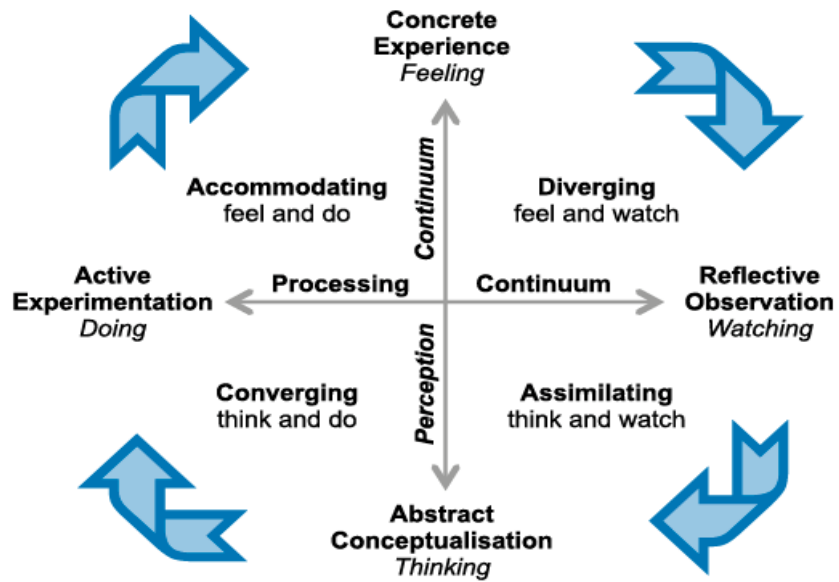

236   RICHARD A. SCHMIDT

Adapted from Patrick (1992)

Appendix C: Dreyfus Model of Skill Acquisition

	Novice	Advanced beginner	Competent	Proficient	Expert
Responsibility extends to others and the environment					Expert
Sense of responsibility increases with experience				Proficient	
Sense of responsibility arises from actively making decisions			Competent		
Still does not experience personal responsibility		Advanced beginner			
Only feels responsible to follow the rules	Novice				

| Scope of vision & Range of capability | Follows specific rules for specific situations. Rules are not conditional. "Only capable of following the rules" | Begins to create and identify conditional rules. "Rules have nuance and become conditional in nature" | Learns organizing principals. Information sorting by relevance begins "Higher order rules shape contexts and conditions" | Uses pattern recognition to assess what to do. Uses rules to determine how to do it "Intuition aides in identifying the situation; the actions are governed by the principals" | No analysis or planning. Pattern recognition extends to plan as well as action "Just does what works" |

Adapted from Kirkpatrick & MacKinnon (2012)

Appendix D: Kolb's Experiential Learning Cycle

Adapted from Van Der Horst & Albertyn (2018)

Appendix E: Contrasting Coaching and Mentoring

	Coaching	Mentoring
1. Level of formality	**More formal**: contract or ground rules set often involving a third party organisational client.	**Less formal**: agreement most typically between two parties.
2. Length of contract	**Shorter term**: typically between 4-12 meetings agreed over two to twelve months.	**Longer term**: typically unspecified number of meetings with relationships often running over 3-5 years.
3. Outcome Focus	**More performance-focused**: typically a greater focus on short term skills and job performance.	**More career-focused**: typically a concern with longer term career issues, obtaining the right experience and longer term thinking.
4. Level of business knowledge	**More generalist**: typically coaches have a strong appreciation of business or commercial realities.	**More sector knowledge**: typically mentors have detailed knowledge of organization or business sector.
5. Training	**More relationship training**: typically coaches have a background in psychology, psychotherapy or human resources, or have undertaken specialist coaching training.	**More management training**: typically mentors have a background in senior management, with limited coaching/mentoring training.
6. Client	**Dual client:** more typically a dual focus on the needs of the individual and the needs of the organisation**.**	**Single client:** more typically a single focus on the needs of the individual**.**
7. Supervision or support	**Formal**: typically the coach will be in (or be expected to be in) supervision as part of their CPD.	**Informal:** typically the mentor may have period discussions or briefings from HR, if based within an organisation.

Adapted from Passmore (2007)

Appendix F: MOSAT Examination

MOSAT examination
Station: vascular anastomosis

Instructions to candidates: One limb of a bifurcated graft has been brought below the inguinal ligament. Perform an end-to-side anastomosis on the common femoral artery.

Item	Not done or done incorrectly	Done correctly
Control of vessel		
1. Appropriate clamp(s) (DeBakey(s) available for the exam)	0	1
2. Clamp(s) placed entirely across artery; may occlude side-to-side or anterior–posterior	0	1
Arteriotomy		
3. Anterior wall, in midline: longitudinal	0	1
4. Uses appropriate blade (#11) to start	0	1
5. Extends with Potts scissors or blade	0	1
6. Appropriate size: 1.5–2 times the circumference of the artery	0	1
7. Avoids trauma to posterior wall of the artery	0	1
Preparing graft		
8. Ensures graft oriented, not twisted	0	1
9. Cuts graft to appropriate length and contour ("cobra-head")	0	1
Anastomosis		
10. Selects appropriate suture (4–0 available for exam)	0	1
11. Selects vascular needle driver and forceps	0	1
12. Starts appropriately: box stitch in heel or parachutes starting few bites from apex	0	1
13. Bites consistently spaced 2–3 mm from one another	0	1
14. Instructs assistant to follow; appropriate tension and direction	0	1
15. Avoids excessive trauma/handling of artery with instruments	0	1
16. Guides suture down to desired position following bites	0	1
17. Ensures that graft lays as patch (on outside) over the arteriotomy	0	1
18. Flushes prior to last few bites	0	1
19. Ties final suture with appropriate tension, no air knot	0	1
20. Minimum of 6 throws on all knots	0	1
21. Needle loaded $1/2$ to $2/3$ >80% of the time	0	1
22. Uses correct needle, needle angle >80% of the time	0	1
23. Bites through artery and graft started at 90° >80% of the time	0	1
24. Follows curve of needle on entrance/exit >80% of the time	0	1
Maximum total score		24
Total score given		

MOSAT = mini objective structured assessment of technical skills

Adapted from Wilasrusmee, Lertsithichai, et al. (2007)

Appendix G: Global Rating Scale

Please rate the candidate's performance on the following scale:							
Respect for tissue	1	2	3	4	5		
	Frequently used unnecessary force on tissues or caused damage by inappropriate instrument use		Careful handling of tissue, but occasional inadvertent damage		Consistently handled tissues appropriately with minimal damage		
Time in motion	1	2	3	4	5		
	Many unnecessary moves		Efficient time and motion, but some unnecessary moves		Clear economy of movement and maximum efficiency		
Instrument handling	1	2	3	4	5		
	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments, but occasionally awkward		Fluid movements		
Suture training	1	2	3	4	5		
	Awkward and unsure with poor knot tying, and inability to maintain tension		Competent suturing with good knot placement and appropriate tension		Excellent suture control with correct suture placement and tension		
Flow of operation	1	2	3	4	5		
	Frequently stopped operating, seemed unsure of next move		Demonstrated some forward planning and reasonable progression of procedure		Obviously planned operation		
Knowledge of procedure	1	2	3	4	5		
	Inefficient knowledge of procedure. Looked unsure and hesitant		Knew all important steps of procedure		Demonstrated familiarity of all steps of procedure		
Final product	1	2	3	4	5		
	Final product of unacceptable quality		Final product of average quality		Final product of superior quality		
Overall performance	1	2	3	4	5		
	Very Poor		Competent		Very Good		
Maximum total score					(40)		
Total score:					( )		
Overall on this task, should the candidate:			Fail		Pass		

Adapted from (Denadai & Souto, 2012)

Appendix H: Informational Waiver of Consent

We would like to ask you to take part in a research study called "Structured, Individualized, Video-Based Coaching to Improve Technical Skill Acquisition." The purpose of this research is to compare self-assessment video analysis against structured, individualized, video-based coaching as a way to improve technical skill acquisition in surgical residents participating in the vascular anastomosis lab.

Your participation in this research project will span the course of the anastomosis lab, which runs for a total of seven weeks. Video recording of your suturing technique will be used throughout the study, which will all be deidentified. Each participant will have full access to only their own deidentified videos throughout the course of the study for practice purposes. In addition to the standard protocol, some participants will engage in one-on-one coaching sessions with the lab's facilitator for a total of three occurrences. After video recording sessions on week one, three, five, and seven, all participants will be asked to evaluate your completed anastomosis using a validated technical skills checklist. This is all that will be asked of you over the course of the seven weeks.

Your taking part in this research is completely voluntary. If you do not wish to participate in the project, you are free to choose not to participate.

Your participation and completion in this clinical research project may not benefit you directly. We are hoping this project will provide information to help us better structure our simulation labs for all our surgical residents who participate in them.

The recorded, deidentified videos and scores gathered from this study will be kept strictly confidential. None of the information you provide will have your name on it that will identify you personally.

If you have any questions about this project or the research study itself, please feel free to ask the Principal Investigator (Kenneth Lynch, NP) to provide you with this information or you may page me at 401-350-2156.

If you have any questions about your rights as a research subject, please feel free to call our Research Protections Office Director, Janice Muratori, at 444-6246.

Thank you very much for your time and assistance with this important educational research.

Kenneth A Lynch, Jr., MS, APRN-CNP, CNSC
Department of Surgery and Critical Care
Rhode Island Hospital
Senior Teaching Associate of Surgery
Alpert Medical School at Brown University

Appendix I: The GROW Coaching Model

Acronym	Description	Example Questions
G – Goal	Coachee is asked to clarify what they want to achieve from each session. Determines the focus of coaching.	What do you want to achieve this session? How would you like to feel afterwards? What would be the best use of this time?
R – Reality	Raise awareness of present realities. Examine how current situation is impacting coachee's goals.	How have things gone in the past week? How have you handled any problems? What worked? What didn't work?
O – Options	Identify and assess available options. Encourage solution-focused thinking and brainstorming.	What possible options do you have? What has worked for you in the past? What haven't you tried yet that might work?
W - Wrap-Up	Assists the coachee determine next steps. Develops an action plan and builds motivation.	What is the most important thing to do next? What might get in the way? Who might be able to support you? How will you feel when this is done?

Adapted from Grant (2011)

Appendix J: The RE-GROW Coaching Model

Adapted from Grant (2011)

Appendix K: Re-GROW Anastomosis Lab Template

Acronym	Description	Example Questions
Review and Evaluate	Session starts by coach reviewing and evaluating the between-session action steps and discuss reflective process	How do you think you did with the anastomosis during this past week? Are you satisfied with your performance?
Goal	Coachee is asked to clarify what they want to achieve from each session. Determines the focus of coaching.	What do you want to work on this week with regards to your performance?
Reality	Raise awareness of present realities. Examine how current situation is impacting coachee's goals.	What do you think you are doing well? How did you do on the scoring evaluation? Where were your deficiencies?
Options	Identify and assess available options. Encourage solution focused thinking and brainstorming.	How would you like to improve on these deficiencies? If you could only work on one aspect of your performance during the next practice session, what would it be?
Wrap-up/Way Forward	Assist the coachee to determine next steps. Develop an action plan and build motivation.	Are there any challenges to this plan moving forward? How can we best work through them?

Adapted from Spence & Grant (2007)

Appendix L: REDCap MOSAT/GRS Electronic Evaluation

Resize font:
➕ | ➖

Mini Objective Structured Assessment of Technical Skills (MOSAT)/Global Rating Scale

Please complete the evaluation below.

1) **Video Session**
   * must provide value

   [▲▼]

2) **Resident ID Number**
   * must provide value

   [▲▼]

3) **Evaluator**
   * must provide value

   [▲▼]

   Mini Objective Structured Assessment of Technical Skills (MOSAT)

   Please Evaluate The Resident's Performance of an End-To-Side Anastomosis:

Preparing Graft

Not Done Correctly = 0
Done Correctly = 1

4)	**Ensures graft oriented, not twisted** * must provide value	○	○ reset
5)	**Cuts end of graft to appropriate length and contour (spatulated)** * must provide value	○	○ reset

Graftotomy

Not Done Correctly = 0
Done Correctly = 1

6)	**Anterior wall, in midline; Longitudinal** * must provide value	○	○ reset
7)	**Uses appropriate blade (#11) to start** * must provide value	○	○ reset
8)	**Extends with Potts scissors or blade** * must provide value	○	○ reset

9)	**Appropriate size; 1cm graftotomy made** * must provide value	○		○	
					reset
10)	**Avoids trauma to the posterior wall of the graft** * must provide value	○		○	
					reset

Anastomosis

Not Done Correctly = 0
Done Correctly = 1

11)	**Selects vascular needle driver and forceps** * must provide value	○		○	
					reset
12)	**Starts appropriately; Knots outside graft** * must provide value	○		○	
					reset
13)	**Correctly uses rubber shod clamp to anchor free needle** * must provide value	○		○	
					reset
14)	**Bites consistently placed, spaced 1-2mm from one another** * must provide value	○		○	
					reset
15)	**Instructs assistant to follow with appropriate tension and direction** * must provide value	○		○	
					reset
16)	**Avoids excessive trauma/handling of graft with instruments** * must provide value	○		○	
					reset
17)	**Guides suture down to desired position following bites** * must provide value	○		○	
					reset
18)	**Ensures that graft lays as a patch (on outside) over the graftotomy** * must provide value	○		○	
					reset
19)	**Inspects anastomosis at regular intervals** * must provide value	○		○	
					reset
20)	**Ties final suture with appropriate tension, no air knot or broken suture** * must provide value	○		○	
					reset
21)	**Minimum of 6 throws on all knots** * must provide value	○		○	
					reset
22)	**Needle loaded 1/2-2/3 > 80% of the time** * must provide value	○		○	
					reset
23)	**Uses correct needle angle > 80% of the time** * must provide value	○		○	
					reset
24)	**Bites through the graft started at 90 degrees > 80% of the time** * must provide value	○		○	
					reset

156

| **25)** | **Follows curve of needle on entrance/exit > 80% of the time**
 * must provide value | ○ | | ○ | |
| | | | | | reset |

Global Rating Scale

Please rate the resident's performance on the following likert scale:

26)	**Respect for Tissue**
	provide value
	requently used unnecessary force on graft or caused damage by inappropriate instrument use
	areful handling of graft, but occasional inadvertent damage
○	onsistently handled graft appropriately with minimal damage

reset

27)	**Time in Motion**
	provide value
	lany unnecessary moves
	fficient time and motion, but some unnecessary moves
○	lear economy of movement and maximum efficiency

reset

28)	**Instrument Handling**
	provide value
	lepeatedly makes tentative or awkward moves with instruments
	ompetent use of instruments, but occasionally awkward
○	luid movements

reset

29)	**Suture training**
	provide value
	wkward and unsure with poor knot tying, and inability to maintain tension
	ompetent suturing with good knot placement and appropriate tension
○	xcellent suture control with correct suture placement and tension

reset

157

30)   **Flow of Operation**

provide value

requently stopped operating, seemed unsure of next move

emonstrated some forward planning and reasonable progression of procedure

○   bviously planned operation

reset

31)   **Knowledge of Procedure**

provide value

nefficient knowledge of procedure. Looked unsure and hesitant

new all important steps of procedure

○   emonstrated familiarity of all steps of procedure

reset

32)   **Final Product**

provide value

inal product of unacceptable quality

inal product of average quality

○   inal product of superior quality

reset

33)   **Overall Performance**

provide value

ery Poor

ompetent

○   xcellent

reset

34)   **Time to complete anastomosis**

35)   **Leak Rate (ml)**

158

Bibliography

Abelson, J. S., Symer, M. M., Yeo, H. L., Butler, P. D., Dolan, P. T., Moo, T. A., & Watkins, A. C. (2018). Surgical time out: our counts are still short on racial diversity in academic surgery. *The American Journal of Surgery*, *215*(4), 542–548.

Accreditation Council for Graduate Medical Education. (2019). *ACGME Program Requirements for Graduate Medical Education in General Surgery*. Chicago, IL.

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior*, *3*(2), 111–150.

Adams, J. A. (1976). Issues for a closed-loop theory of motor learning. In *Motor Control* (pp. 87–107). Elsevier.

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin*, *101*(1), 41.

Adesunloye, B. A., Aladesanmi, O., Henriques-Forsythe, M., & Ivonye, C. (2008). The preferred learning style among residents and faculty members of an internal medicine residency program. *Journal of the National Medical Association*, *100*(2), 172–177.

Agha, R. A., & Fowler, A. J. (2015). The role and validity of surgical simulation. *International Surgery*, *100*(2), 350–357.

Ahlberg, G., Enochsson, L., Gallagher, A. G., Hedman, L., Hogman, C., McClusky III, D. A., … Arvidsson, D. (2007). Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American Journal of Surgery*, *193*(6), 797–804.

Altman, D. G. (1985). Comparability of randomised groups. *Journal of the Royal

Statistical Society: Series D (The Statistician), *34*(1), 125–136.

Alvand, A., Auplish, S., Gill, H., & Rees, J. (2011). Innate arthroscopic skills in medical students and variation in learning curves. *JBJS*, *93*(19), e115.

Anderson, D. I., Magill, R. A., & Sekiya, H. (2001). Motor learning as a function of KR schedule and characteristics of task-intrinsic feedback. *Journal of Motor Behavior*, *33*(1), 59–66.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science*, *13*(4), 467–505.

Arnold, L., Willoughby, T. L., & Calkins, E. (1985). Self-evaluation in undergraduate medical education: a longitudinal perspective. *Journal of Medical Education*.

Arora, S., Miskovic, D., Hull, L., Moorthy, K., Aggarwal, R., Johannsson, H., … Sevdalis, N. (2011). Self vs expert assessment of technical and non-technical skills in high fidelity simulation. *The American Journal of Surgery*, *202*(4), 500–506.

Atkins, J. L., Kalu, P. U., Lannon, D. A., Green, C. J., & Butler, P. E. (2005). Training in microsurgical skills: does course-based learning deliver? *Microsurgery*, *25*(6), 481–485.

Backstein, D., Agnidis, Z., Regehr, G., & Reznick, R. (2004). The effectiveness of video feedback in the acquisition of orthopedic technical skills. *American Journal of Surgery*, *187*(3), 427–432. https://doi.org/10.1016/j.amjsurg.2003.12.011

Backstein, D., Agnidis, Z., Sadhu, R., & MacRae, H. (2005). Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. *Canadian Journal of Surgery. Journal Canadien de Chirurgie*, *48*(3), 195–200. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16013622

Baker, J. D. (1987). Bleeding through PTFE grafts. *European Journal of Vascular Surgery*, *1*(1), 41–43.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Barsuk, J. H., Cohen, E. R., Williams, M. V., Scher, J., Jones, S. F., Feinglass, J., … Wayne, D. B. (2018). Simulation-Based Mastery Learning for Thoracentesis Skills Improves Patient Outcomes: A Randomized Trial. *Academic Medicine*, *93*(5), 729–735. https://doi.org/10.1097/ACM.0000000000001965

Benner, P. (2001). From novice to expert (Commemorative Edition). In *Pretense Hall Health*. Upper Saddle River.

Benner, Patricia. (1982). From novice to expert. *AJN The American Journal of Nursing*, *82*(3), 402–407.

Benner, Patricia. (2004). Using the Dreyfus model of skill acquisition to describe and interpret skill acquisition and clinical judgment in nursing practice and education. *Bulletin of Science, Technology & Society*, *24*(3), 188–199.

Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, *24*(3), 200–212.

Bevan, P. G. (1981). *THE CRAFT OF SURGERY-THE ANASTOMOSIS WORKSHOP, MARCH 1981*. ROYAL COLL SURGEONS ENGLAND 35-43 LINCOLN'S INN FIELDS, LONDON, ENGLAND WC2A 3PN.

Bishop, J. (2015). An investigation into the extent and limitations of the GROW model for coaching and mentoring online: Towards' prosthetic learning'. *Proceedings of the International Conference on E-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*, 125. The Steering Committee of The World

Congress in Computer Science, Computer ....

Bloom, B. S., & Sosniak, L. A. (1985). *Developing talent in young people*. Ballantine

Books.

Bodemer, C. W. (1983). France, the fundament, and the rise of surgery. *Diseases of the

Colon & Rectum*, *26*(11), 743–750. https://doi.org/10.1007/BF02554989

Bolster, L., & Rourke, L. (2015, September 1). The Effect of Restricting Residents' Duty

Hours on Patient Safety, Resident Well-Being, and Resident Education: An Updated

Systematic Review. *Journal of Graduate Medical Education*, Vol. 7, pp. 349–363.

https://doi.org/10.4300/JGME-D-14-00612.1

Bonrath, E. M., Dedy, N. J., Gordon, L. E., & Grantcharov, T. P. (2015). Comprehensive

surgical coaching enhances surgical skill in the operating room: A randomized

controlled trial. *Annals of Surgery*, *262*(2), 205–212.

https://doi.org/10.1097/SLA.0000000000001214

Book, W. F. (1925). *The Psychology of Skill New York* (Gregg, Ed.). New York.

Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self assessment

and reflection in professional courses. *Social Work Education*, *18*(2), 121–132.

Boud, D. (2013). *Enhancing learning through self-assessment*. Routledge.

Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning experience into learning*.

Routledge.

Branscheidt, M., Kassavetis, P., Anaya, M., Rogers, D., Huang, H. D., Lindquist, M. A.,

& Celnik, P. (2019). Fatigue induces long-lasting detrimental changes in motor-skill

learning. *ELife*, *8*, e40578.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor

memory. *Nature*, *382*, 252–255.

Brennan, T. A., Leape, L. L., Laird, N. M., Hebert, L., Localio, A. R., Lawthers, A. G., … Hiatt, H. H. (1991). Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *The New England Journal of Medicine*, *324*(6), 370–376. https://doi.org/10.1056/NEJM199102073240604

Bruce, A. N., Battista, A., Plankey, M. W., Johnson, L. B., & Marshall, M. B. (2015). Perceptions of gender-based discrimination during surgical training and practice. *Medical Education Online*, *20*(1), 25923.

Bryan, R. E., Krych, A. J., Carmichael, S. W., Viggiano, T. R., & Pawlina, W. (2005). Assessing professionalism in early medical education: experience with peer evaluation and self-evaluation in the gross anatomy course. *Annals-Academy of Medicine Singapore*, *34*(8), 486–491.

Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, *4*(1), 27–53.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, *6*(4), 345–375.

Brydges, R., Hatala, R., Zendejas, B., Erwin, P. J., & Cook, D. A. (2015, February 6). Linking simulation-based educational assessments and patient-related outcomes: A systematic review and meta-analysis. *Academic Medicine*, Vol. 90, pp. 246–256. https://doi.org/10.1097/ACM.0000000000000549

Cameron, J. L. (1997). William Stewart Halsted. Our surgical heritage. *Annals of Surgery*, *225*(5), 445–458. https://doi.org/10.1097/00000658-199705000-00002

Carraccio, C. L., Benson, B. J., Nixon, L. J., & Derstine, P. L. (2008). From the

educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. *Academic Medicine*, *83*(8), 761–767.

Cass, O. W. (1999). Training to competence in gastrointestinal endoscopy: a plea for continuous measuring of objective end points. *Endoscopy*, *31*(09), 751–754.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Cogbill, T. H., Malangoni, M. A., Potts, J. R., & Valentine, R. J. (2014). The general surgery milestones project. *Journal of the American College of Surgeons*, *218*(5), 1056–1062. https://doi.org/10.1016/j.jamcollsurg.2014.02.016

Coleman, J. J., Esposito, T. J., Rozycki, G. S., & Feliciano, D. V. (2013). Early subspecialization and perceived competence in surgical training: Are residents ready? *Journal of the American College of Surgeons*, *216*(4), 764–771. https://doi.org/10.1016/j.jamcollsurg.2012.12.045

Collins, A., Brown, J. S., & Newman, S. (1989). *Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. Knowing, learning, and instruction: Essays in honor of Robert Glaser*.

Colthart, I., Bagnall, G., Evans, A., Allbutt, H., Haig, A., Illing, J., & McKinstry, B. (2008). The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. *Medical Teacher*, *30*(2), 124–145.

Cope, Z. (1959). *The Royal College of Surgeons of England: A History* (A. Blonde, Ed.). London.

Creswell, J. W., & Creswell, J. D. (2014). *Research design: Qualitative, quantitative, and*

mixed methods approaches. Sage publications.

Dang, M.-C., Thacker, J. G., Hwang, J. C. S., Rodeheaver, G. T., Melton, S. M., & Edlich, R. F. (1990). Some biomechanical considerations of polytetrafluoroethylene sutures. *Archives of Surgery*, *125*(5), 647–650.

Datta, V., Mandalia, M., Mackay, S., Chang, A., Cheshire, N., & Darzi, A. (2002). Relationship between skill and outcome in the laboratory-based model. *Surgery*, *131*(3), 318–323.

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Jama*, *296*(9), 1094–1102.

Dawe, S. R., Pena, G. N., Windsor, J. A., Broeders, J. A. J. L., Cregan, P. C., Hewett, P. J., & Maddern, G. J. (2014). Systematic review of skills transfer after surgical simulation-based training. *British Journal of Surgery*, Vol. 101, pp. 1063–1076. https://doi.org/10.1002/bjs.9482

Denadai, R., & Souto, L. R. M. (2012). Organic bench model to complement the teaching and learning on basic surgical skills. *Acta Cirurgica Brasileira*, *27*(1), 88–94.

Dewey, J. (1938). *Education and Experience* (Simon and Schuster, Ed.). New York.

Donaldson, M. S., Corrigan, J. M., & Kohn, L. T. (1999). Institute of Medicine. To Err Is Human: Building a Safer Health System. *To Err Is Human: Building a Safer Health System.*, *6*. https://doi.org/10.1017/S095026880100509X

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*(5), 795.

Dreyfus, H. L., & Dreyfus, S. E. (1986). The power of human intuition and expertise in

the era of the computer. In *Mind over machine*. New York: The Free Press.

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science,*

Technology & Society, *24*(3), 177–181.

Dreyfus, S. E., & Dreyfus, H. L. (1979). *The scope, limits, and training implications of*

three models of aircraft pilot emergency response behavior. CALIFORNIA UNIV

BERKELEY OPERATIONS RESEARCH CENTER.

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A FIVE-STAGE MODEL OF THE MENTAL*

ACtITIES INVOLVED IN DIRECTED SKILL ACQUISITION OPERATIONS " '-fj-

RESEARCH.

Dunnington, G. L. (1996). The art of mentoring. *The American Journal of Surgery*,

171(6), 604–607. https://doi.org/10.1016/S0002-9610(95)00028-3

Ebbinghaus, H. (1964). *Memory: Acontribution to experimental psychology*. Dover.

Elkins, M. R. (2015). Assessing baseline comparability in randomised trials. *Journal of*

Physiotherapy, *61*(4), 228–230.

Engels, P. T., & de Gara, C. (2010). Learning styles of medical students, general surgery

residents, and general surgeons: implications for surgical education. *BMC Medical*

Education, *10*(1), 51.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert

performance in medicine and related domains. *Academic Medicine*, *79*(10), S70–

S81.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the

development of superior expert performance. *The Cambridge Handbook of*

Expertise and Expert Performance, *38*, 685–705.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate

practice in the acquisition of expert performance. *Psychological Review*, *100*(3),

363.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance:

Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*,

47(1), 273–305.

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2009). Toward a science of

exceptional achievement. *Annals of the New York Academy of Sciences*, *1172*(1),

199.

Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: a

reformulation and research agenda. *Academic Medicine*, *80*(10), S46–S54.

Eva, K. W., & Regehr, G. (2007). Knowing when to look it up: A new conception of self-

assessment ability. *Academic Medicine*, *82*(10 SUPPL.), s81–s84.

https://doi.org/10.1097/ACM.0b013e31813e6755

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-

analysis. *Review of Educational Research*, *59*(4), 395–430.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A

meta-analysis comparing peer and teacher marks. *Review of Educational Research*,

70(3), 287–322.

Farquharson, A. L., Cresswell, A. C., Beard, J. D., & Chan, P. (2013). Randomized trial

of the effect of video feedback on the acquisition of surgical skills. *British Journal

of Surgery*, *100*(11), 1448–1453.

Fee, E. (2015). The first American medical school: The formative years. *The Lancet*, *385*(9981), 1940–1941. https://doi.org/10.1016/S0140-6736(15)60950-3

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education*. New York: McGraw-Hill Humanities/Social Sciences/Languages.

Franzese, C. B., & Stringer, S. P. (2007, December). The Evolution of Surgical Training: Perspectives on Educational Models from the Past to the Future. *Otolaryngologic Clinics of North America*, Vol. 40, pp. 1227–1235. https://doi.org/10.1016/j.otc.2007.07.004

Fronza, J. S., Prystowsky, J. P., Darosa, D., & Fryer, J. P. (2012). Surgical residents' perception of competence and relevance of the clinical curriculum to future practice. *Journal of Surgical Education*, *69*(6), 792–797. https://doi.org/10.1016/j.jsurg.2012.05.014

Gaba, D. M. (2004). The future vision of simulation in health care. *BMJ Quality & Safety*, *13*(suppl 1), i2–i10.

Gallagher, A. G., Jordan-Black, J. A., & O'Sullivan, G. C. (2012). Prospective, randomized assessment of the acquisition, maintenance, and loss of laparoscopic skills. *Annals of Surgery*, *256*(2), 387–393. https://doi.org/10.1097/SLA.0b013e318251f3d2

Gallwey, W. T. (2002). The Inner Game of Work: Overcoming Mental Obstacles for Maximum Performance. *New York: Texere*.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences* (Vol. 27). Macmillan.

George, B. C., Bohnen, J. D., Williams, R. G., Meyerson, S. L., Schuller, M. C., Clark, M. J., … Fryer, J. P. (2017). Readiness of US General Surgery Residents for Independent Practice. *Annals of Surgery*, *266*(4), 582–594. https://doi.org/10.1097/SLA.0000000000002414

Gill, H. (1972). *The apothecary in colonial Virginia*. Retrieved from https://books.google.com/books?hl=en&lr=&id=0-LhEY6IjyUC&oi=fnd&pg=PP9&dq=The+apothecary+in+colonial+Virginia&ots=hii01hpYup&sig=DPvK3p46ZjLA5qTpIGChauVReW8

Gladwell, M. (2008). *Outliers: The story of success*. Hachette UK.

Goldowsky, S. (1988). Yankee Surgeon: The Life and Times of Usher Parsons, 1788-1868. In *Surgery*. Watson Pub International.

Gordon, M. J. (1991). A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine*.

Grant, A. M. (2001). Towards a psychology of coaching. *Unpublished Manuscript, Sydney*.

Grant, A. M. (2011). Is it time to REGROW the GROW model? Issues related to teaching coaching session structures. *The Coaching Psychologist*, *7*(2), 118–126.

Grantcharov, T. P., & Funch-Jensen, P. (2009). Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. *The American Journal of Surgery*, *197*(4), 447–449.

Greenhalgh, R. M., & Flack, I. (1981). The buckston browne arterial jig. *Ann Roy Coll Surg Engl*, *63*, 408–410.

Grillo, H. C. (1999). To impart this art: The development of graduate surgical education

in the United States. *Surgery*, *125*(1), 1–14. https://doi.org/10.1016/S0039-
6060(99)70280-1

Grillo, H. C. (2004). Edward D. Churchill and the "rectangular" surgical residency.
Surgery, *136*(5), 947–952. https://doi.org/10.1016/j.surg.2004.09.002

Hall, J. C. (2002). Imagery practice and the development of surgical skills. *The American
Journal of Surgery*, *184*(5), 465–470.

Halm, E. A., Lee, C., & Chassin, M. R. (2002). Is volume related to outcome in health
care? A systematic review and methodologic critique of the literature. *Annals of
Internal Medicine*, *137*(6), 511–520.

Halsted, W. S. (1904). The training of the surgeon. *Bull Johns Hop Hosp*, (xv), 267–275.

Hamdorf, J. M., & Hall, J. C. (2000). Acquiring surgical skills. *British Journal of
Surgery*, *87*(1), 28–37. https://doi.org/10.1046/j.1365-2168.2000.01327.x

Hargreaves, J., & Lane, D. (2001). Delya's story: from expert to novice, a critique of
Benner's concept of context in the development of expert nursing practice.
International Journal of Nursing Studies, *38*(4), 389–394.

Hashimoto, D. A., Sirimanna, P., Gomez, E. D., Beyer-Berjot, L., Ericsson, K. A.,
Williams, N. N., … Aggarwal, R. (2015). Deliberate practice enhances quality of
laparoscopic surgical performance in a randomized controlled trial: from arrested
development to expert performance. *Surgical Endoscopy*, *29*(11), 3154–3162.

Hatala, R., Cook, D. A., Brydges, R., & Hawkins, R. (2015). Constructing a validity
argument for the Objective Structured Assessment of Technical Skills (OSATS): a
systematic review of validity evidence. *Advances in Health Sciences Education*,
20(5), 1149–1175.

Hawkins, S. C., Osborne, A., Schofield, S. J., Pournaras, D. J., & Chester, J. F. (2012). Improving the accuracy of self-assessment of practical clinical skills using video feedback–the importance of including benchmarks. *Medical Teacher*, *34*(4), 279–284.

Head, H. (1925). *Aphasia and kindred disorders of speech*. Cambridge University Press.

Helling, T. (2016). *Desperate Surgery in the Pacific War: Doctors and Damage Control for American Wounded, 1941–1945*. Retrieved from https://books.google.com/books?hl=en&lr=&id=ARwCDgAAQBAJ&oi=fnd&pg=PP1&dq=Desperate+Surgery+in+the+Pacific+War:+Doctors+and+Damage+Control+for+&ots=XRky-2z2Gd&sig=64ECmCbHODFTSMm_fpDlZNtJXTM

Herrera-Almario, G. E., Kirk, K., Guerrero, V. T., Jeong, K., Kim, S., & Hamad, G. G. (2016). The effect of video review of resident laparoscopic surgical skills measured by self- and external assessment. *American Journal of Surgery*, *211*(2), 315–320. https://doi.org/10.1016/j.amjsurg.2015.05.039

Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 4). John Wiley & Sons.

Hu, Y., Tiemann, D., & Brunt, M. (2013). Video self-assessment of basic suturing and knot tying skills by novice trainees. *Journal of Surgical Education*, *70*(2), 279–283. https://doi.org/10.1016/j.jsurg.2012.10.003

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory.*

Jensen, A. R., Milner, R., Achildi, O., Gaughan, J., Wilhite, D. B., & Grewal, H. (2008). Effective instruction of vascular anastomosis in the surgical skills laboratory. *American Journal of Surgery*, *195*(2), 189–194.

https://doi.org/10.1016/j.amjsurg.2007.09.032

Kahol, K., Leyba, M. J., Deka, M., Deka, V., Mayes, S., Smith, M., … Panchanathan, S. (2008). Effect of fatigue on psychomotor and cognitive skills. *The American Journal of Surgery*, *195*(2), 195–204.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science*, *265*(5172), 679–682.

Kaufman, M. (1976). *American Medical Education: The Formative Years, 1765-1910*. Westport, CT: Greenwood Press.

Kaushik, P. (2017). Redefining Learning: Kolb's Theory of Learning Styles with Gardner's. *International Journal of Learning and Teaching*, *9*(1), 330–339.

King, L. S. (1983). XVII. Medical Education: Elitisms and Reform. *JAMA: The Journal of the American Medical Association*, *250*(18), 2457–2461. Retrieved from https://jamanetwork.com/

Kirkpatrick, K., & MacKinnon, R. J. (2012). Technology-enhanced learning in anaesthesia and educational theory. *Continuing Education in Anaesthesia, Critical Care & Pain*, *12*(5), 263–267.

Kneebone, R., Arora, S., King, D., Bello, F., Sevdalis, N., Kassab, E., … Nestel, D. (2010). Distributed simulation–accessible immersive training. *Medical Teacher*, *32*(1), 65–70.

Knowles, M. S. (1980). *The modern practice of adult education*.

Kolb, A., & Kolb, D. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning and Education*,

4(2), 193–212. https://doi.org/10.5465/AMLE.2005.17268566

Kolb, D. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2001). Experiential learning theory: Previous research and new directions. *Perspectives on Thinking, Learning, and Cognitive Styles*, *1*(8), 227–247.

Korndorffer Jr, J. R., Dunne, J. B., Sierra, R., Stefanidis, D., Touchard, C. L., & Scott, D. J. (2005). Simulator training for laparoscopic suturing using performance goals translates to the operating room. *Journal of the American College of Surgeons*, *201*(1), 23–29.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 years of consolidation—remembering Müller and Pilzecker. *Learning & Memory*, *6*(2), 77–87.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, *59*(4), 277–287.

Lerner, B. (2006). A case that shook medicine. *The Washington Post*. Retrieved from http://cores33webs.mede.uic.edu/uicpatientsafetyeducation/medstudentcurriculum/e dresources/documents/UIC_patient_safety_curriculums/UIC_M1-M3_curriculum/Stress in Medicine/M1_activities/A_Case_That_Shook_Medicine_article.pdf

Lewin, K. (1935). *A dynamic theory of personality* (McGraw-Hill, Ed.). New York.

Limberg, A. A. (1984). *The planning of local plastic operations on the body surface: theory and practice*. Collamore Press.

Lodge, D., & Grantcharov, T. (2011). Training and assessment of technical skills and competency in cardiac surgery. *European Journal of Cardio-Thoracic Surgery*, *39*(3), 287–293.

Loudon, I. (2000). Why are (male) surgeons still addressed as Mr? *BMJ*, *321*(7276), 1589. https://doi.org/10.1136/bmj.321.7276.1589

Louie, K., & Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, *29*(1), 145–156. https://doi.org/10.1016/S0896-6273(01)00186-6

Lounsbery, M. F., & Sharpe, T. (1996). Plotting the effects of corrective self-analysis on elite volleyball skill performance. *Applied Research in Coaching and Athletics Annual*, 31–42.

Lyon, L. J. (2015). Development of teaching expertise viewed through the Dreyfus model of skill acquisition. *Journal of the Scholarship of Teaching and Learning*, 88–105.

Mammen, J. M. V, Fischer, D. R., Anderson, A., James, L. E., Nussbaum, M. S., Bower, R. H., & Pritts, T. A. (2007). Learning styles vary among general surgery residents: analysis of 12 years of data. *Journal of Surgical Education*, *64*(6), 386–389.

Martin, D., Regehr, G., Hodges, B., & McNaughton, N. (1998). Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Academic Medicine*, *73*(11), 1201–1206.

Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C., &

Brown, M. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery*, *84*(2), 273–278.

Mattar, S. G., Alseidi, A. A., Jones, D. B., Jeyarajah, D. R., Swanstrom, L. L., Aye, R. W., … Minter, R. M. (2013). General surgery residency inadequately prepares trainees for fellowship: Results of a survey of fellowship program directors. *Annals of Surgery*, *258*(3), 440–447. https://doi.org/10.1097/SLA.0b013e3182a191ca

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, *287*(5451), 248–251.

Mellinger, J. D., Williams, R. G., Sanfey, H., Fryer, J. P., DaRosa, D., George, B. C., … Scott, D. J. (2017). Teaching and assessing operative skills: From theory to practice. *Current Problems in Surgery*, *54*(2), 44–81. https://doi.org/10.1067/j.cpsurg.2016.11.007

Merriam-Webster Dictionary. (n.d.). Merriam-Webster Dictionary. Retrieved from Merriam-Webster Dictionary website: https://www.merriam-webster.com/dictionary/competence

Merriam, S., Caffarella, R., & Baumgartner, L. (2007). *Learning in adulthood: A comprehensive guide*. Retrieved from https://books.google.com/books?hl=en&lr=&id=5o1UAwAAQBAJ&oi=fnd&pg=PR9&dq=Learning+in+Adulthood:+A+Comprehensive+Guide%22&ots=vqNtzRe8dX&sig=ij4CKFfDdhpUMkQYUTbsMGY2FaA

Mery, C. M., Greenberg, J. A., Patel, A., & Jaik, N. P. (2008). Teaching and assessing the ACGME competencies in surgical residency. *Bulletin of the American College of Surgeons*, *93*(7), 39–47. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/19469344

Miller, L. E., & Weiss, R. (2008). Medical Education Reform Efforts and Failures of U.S. Medical Schools, 1870-1930. *Journal of the History of Medicine and Allied Sciences*, *63*(3), 348–387. https://doi.org/10.1093/jhmas/jrn002

Min, H., Morales, D. R., Orgill, D., Smink, D. S., & Yule, S. (2015, November 1). Systematic review of coaching to enhance surgeons' operative performance. *Surgery (United States)*, Vol. 158, pp. 1168–1191. https://doi.org/10.1016/j.surg.2015.03.007

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., … Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology*, *63*(8), e1–e37.

Moore, M. J., & Bennett, C. L. (1995). The learning curve for laparoscopic cholecystectomy. *The American Journal of Surgery*, *170*(1), 55–59.

Morgan, C. N. (1968). Surgery and surgeons in 18th-century London. *Annals of the Royal College of Surgeons of England*, *42*(1), 1–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4865018

Mory, E. H. (2004). Feedback research revisited. *Handbook of Research on Educational Communications and Technology*, *2*, 745–783.

Moulton, C.-A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Annals of Surgery*, *244*(3), 400.

Müller, G. E., & Pilzecker, A. (1900). *Experimentelle beiträge zur lehre vom gedächtniss* (Vol. 1). JA Barth.

Nagendran, M., Gurusamy, K. S., Aggarwal, R., Loizidou, M., & Davidson, B. R. (2013, August 27). Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database of Systematic Reviews*, Vol. 2013. https://doi.org/10.1002/14651858.CD006575.pub3

Nasca, T. J., Day, S. H., & Amis, E. S. (2010). The new recommendations on duty hours from the ACGME task force. *New England Journal of Medicine*, *363*(2), e3(1)-e3(6). https://doi.org/10.1056/NEJMsb1005800

Okhah, Z., Morrissey, P., Harrington, D. T., Cioffi, W. G., & Charpentier, K. P. (2013). Assessment of surgical residents in a vascular anastomosis laboratory. *Journal of Surgical Research*, *185*(1), 450–454.

Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. *Handbook of Mixed Methods in Social and Behavioral Research*, *2*, 397–430.

Oxford English Dictionary. (2017). Oxford English Dictionary Online.

Palter, V. N., & Grantcharov, T. P. (2014). Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: a randomized controlled trial. *Annals of Surgery*, *259*(3), 443–448.

Pandey, V. A., Wolfe, J. H. N., Black, S. A., Cairols, M., Liapis, C. D., & Bergqvist, D. (2008). Self-assessment of technical skill in surgery: the need for expert feedback. *The Annals of The Royal College of Surgeons of England*, *90*(4), 286–290.

Parker, R. W., Alford, C., & Passmore, C. (2004). Can family medicine residents predict their performance on the in-training examination? *FAMILY MEDICINE-KANSAS CITY-*, *36*(10), 705–709.

Passmore, J. (2007). Coaching and mentoring: The role of experience and sector knowledge. *International Journal of Evidence Based Coaching and Mentoring*, *5*(S1), 10–16.

Patel, N., & Popp, R. L. (2014). Learning Lessons. *Journal of the American College of Cardiology*, *64*(25), 2802–2804.

Patrick, J. (1992). *Training: Research and Practice*. Academic Press, London, UK.

Peña, A. (2010). The Dreyfus model of clinical problem-solving skills acquisition: a critical perspective. *Medical Education Online*, *15*(1), 1–11.

Peno, K., & Silva Mangiante, E. (2012). The journey from novice to expert: Toward a purposeful on-going mentoring model. In *C. Boden, & K. King, Conversations about Adult Learning in a Complex World. Charlotte, NC: Information Age Publishing*. Charlotte, NC: Information Age Publishing.

Philibert, I., Friedmann, P., & Williams, W. T. (2002, September 4). New requirements for resident duty hours. *Journal of the American Medical Association*, Vol. 288, pp. 1112–1114. https://doi.org/10.1001/jama.288.9.1112

Piaget, J. (1980). *Experiments in contradiction*. University of Chicago Press.

Polavarapu, H., Kulaylat, A., Sun, S., & Hamed, O. (2013). 100 years of surgical education: the past, present, and future. *Bulletin.Facs.Org*, *98*(7), 22–27. Retrieved from http://bulletin.facs.org/2013/07/100-years-of-surgical-education/

Purkis, M. E. (1994). Entering the field: intrusions of the social and its exclusion from studies of nursing practice. *International Journal of Nursing Studies*, *31*(4), 315–336.

Reznick, R., Regehr, G., MacRae, H., Martin, J., & McCulloch, W. (1997). Testing

technical skill via an innovative "bench station" examination. *The American Journal of Surgery*, *173*(3), 226–230.

Risucci, D. A., Tortolani, A. J., & Ward, R. J. (1989). Ratings of surgical residents by self, supervisors and peers. *Surgery, Gynecology & Obstetrics*, *169*(6), 519–526.

Ritchie, W. P. (2001). The measurement of competence. Current plans and future initiatives of the American Board of Surgery. *Bulletin of the American College of Surgeons*, *86*(4), 10–15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11357767

Rizan, C., Ansell, J., Tilston, T. W., Warren, N., & Torkington, J. (2015). Are general surgeons able to accurately self-assess their level of technical skills? *The Annals of The Royal College of Surgeons of England*, *97*(8), 549–555.

Roberts, C., & Torgerson, D. J. (1999). Baseline imbalance in randomised controlled trials. *Bmj*, *319*(7203), 185.

Robinson, J. O. (1984). The Barber-Surgeons of London. *Archives of Surgery*, *119*(10), 1171–1175. https://doi.org/10.1001/archsurg.1984.01390220053012

Rogers, A. C., Turley, L. P., Cross, K. S., & McMonagle, M. P. (2016). Meta□analysis of the use of surgical sealants for suture□hole bleeding in arterial anastomoses. *British Journal of Surgery*, *103*(13), 1758–1767.

Rogers, H., & Münsterberg, H. (1906). *Congress of arts and science, Universal exposition, St. Louis, 1904* (M. and company. Houghton, Ed.). Retrieved from https://books.google.com/books?hl=en&lr=&id=D_UpAQAAIAAJ&oi=fnd&pg=PA3&dq=congress+of+arts+and+science+1906&ots=7f8SDLKXya&sig=aVYfqO9LgT-F2OZiPJI9bwtjsGM

Rosen, J. M., Long, S. A., McGrath, D. M., & Greer, S. E. (2009). Simulation in plastic

    surgery training and education: the path forward. *Plastic and Reconstructive*

    Surgery, *123*(2), 729–738. https://doi.org/10.1097/PRS.0b013e3181958ec4

Ross, J. A. (2006). *The reliability, validity, and utility of self-assessment*.

Rudge, T. (1992). Reflections on Bennet: a critical perspective. *Contemporary Nurse*,

    1(2), 84–88.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor

    learning: a review and critical reappraisal. *Psychological Bulletin*, *95*(3), 355–386.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological*

    Review, *82*(4), 225.

Schmidt, R. A. (2003). Motor schema theory after 27 years: Reflections and implications

    for a new theory. *Research Quarterly for Exercise and Sport*, *74*(4), 366–375.

Schmidt, R. A., Lee, T. D., Winstein, C., Wulf, G., & Zelaznik, H. N. (2018). *Motor*

    control and learning: A behavioral emphasis. Human kinetics.

Schmidt, R. A., & Wrisberg, C. A. (2004). *Motor learning and performance, A problem-*

    based approach. Champaign, Illinois: Human Kinetics.

Schon, D. (1983). *The Reflective Practitioner New York*. New York: Harper & Collins.

Schön, D. A. (1987). *Educating the reflective practitioner*.

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*.

    Routledge.

Schwartz, S., de Virgilio, M., Chisum, P., Heslin, A., Zein, A., McDermott, M., … de

    Virgilio, C. (2014). A Prospective Randomized Study Assessing Optimal Method

    for Teaching Vascular Anastomoses. *Annals of Vascular Surgery*, *28*(5), 1087–

1093.

Schwartz, S. I., Galante, J., Kaji, A., Dolich, M., Easter, D., Melcher, M. L., … De

Virgilio, C. (2013). Effect of the 16-hour work limit on general surgery intern

operative case volume: A multi-institutional study. *JAMA Surgery*, *148*(9), 829–833.

https://doi.org/10.1001/jamasurg.2013.2677

Sen, S., Kranzler, H. R., Didwania, A. K., Schwartz, A. C., Amarnath, S., Kolars, J. C.,

… Guille, C. (2013). Effects of the 2011 duty hour reforms on interns and their

patients: A prospective longitudinal cohort study. *JAMA Internal Medicine*, *173*(8),

657–662. https://doi.org/10.1001/jamainternmed.2013.351

Shea, C. H., Lai, Q., Black, C., & Park, J.-H. (2000). Spacing practice sessions across

days benefits the learning of motor skills. *Human Movement Science*, *19*(5), 737–

760.

Shryock, R. H. (1965). European Backgrounds of American Medical Education (1600-

1900). *JAMA: The Journal of the American Medical Association*, *194*(7), 709–714.

Retrieved from https://jamanetwork.com/

Sidhu, R. S., Park, J., Brydges, R., MacRae, H. M., & Dubrowski, A. (2007). Laboratory-

based vascular anastomosis training: a randomized controlled trial evaluating the

effects of bench model fidelity and level of training on skill acquisition. *Journal of

Vascular Surgery*, *45*(2), 343–349.

Singh, P., Aggarwal, R., Tahir, M., Pucher, P. H., & Darzi, A. (2015). A Randomized

Controlled Study to Evaluate the Role of Video-based Coaching in Training

Laparoscopic Skills. *Annals of Surgery*, *261*(5), 862–869.

https://doi.org/10.1097/SLA.0000000000000857

Skinner, B. (1965). F.(1953). Science and human behavior. *New York*.

Soucisse, M. L., Boulva, K., Sideris, L., Drolet, P., Morin, M., & Dubé, P. (2017). Video

Coaching as an Efficient Teaching Method for Surgical Residents—A Randomized

Controlled Trial. *Journal of Surgical Education*, *74*(2), 365–371.

https://doi.org/10.1016/j.jsurg.2016.09.002

Spence, G. B., & Grant, A. M. (2007). Professional and peer life coaching and the

enhancement of goal striving and well-being: An exploratory study. *The Journal of*

Positive Psychology, *2*(3), 185–194.

Stain, S. C. (2015). Contributions of public hospitals to surgery in the United States.

Journal of the American College of Surgeons, *221*(1), 1–6.

https://doi.org/10.1016/j.jamcollsurg.2014.12.060

Stefanidis, D., & Heniford, B. T. (2009). The formula for a successful laparoscopic skills

curriculum. *Archives of Surgery*, *144*(1), 77–82.

Stefanidis, D., Korndorffer Jr, J. R., Heniford, B. T., & Scott, D. J. (2007). Limited

feedback and video tutorials optimize learning and resource utilization during

laparoscopic simulator training. *Surgery*, *142*(2), 202–206.

Stefanidis, D., Korndorffer Jr, J. R., Sierra, R., Touchard, C., Dunne, J. B., & Scott, D. J.

(2005). Skill retention following proficiency-based laparoscopic simulator training.

Surgery, *138*(2), 165–170.

Stickgold, R., James, L., & Hobson, J. A. (2000). Visual discrimination learning requires

sleep after training. *Nature Neuroscience*, *3*(12), 1237.

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., & Hobson, J. A. (2000). Visual

discrimination task improvement: A multi-step process occurring during sleep.

Journal of Cognitive Neuroscience, *12*(2), 246–254.

Street, H. (2002, July). Exploring relationships between goal setting, goal pursuit and depression: A review. *Australian Psychologist*, Vol. 37, pp. 95–103. https://doi.org/10.1080/00050060210001706736

Strowd, R. E., & Reynolds, P. (2013). Opinion & Special Articles: The lost resident: Why resident physicians still need mentoring. *Neurology*, *80*(23), e244–e246.

Swing, S. R. (2007). The ACGME outcome project: retrospective and prospective. *Jama*, *288*(9), 648–654. https://doi.org/10.1080/01421590701392903

Szasz, P., Louridas, M., Harris, K. A., Aggarwal, R., & Grantcharov, T. P. (2015, June 1). Assessing technical competence in surgical trainees: A systematic review. *Annals of Surgery*, Vol. 261, pp. 1046–1055. https://doi.org/10.1097/SLA.0000000000000866

Temkin, O. (1951). The role of surgery in the rise of modern medical thought. *Bulletin of the History of Medicine*, *24*, 248–259. Retrieved from http://search.proquest.com/openview/b7f8cd9483ca54fa848920285a0fec05/1?pq-origsite=gscholar&cbl=1816629

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, *39*, 212–222. https://doi.org/10.2307/1415413

Tsuda, S., Scott, D., Doyle, J., & Jones, D. B. (2009). Surgical skills training and simulation. *Current Problems in Surgery*, *4*(46), 271–370.

Tyte, K. (2011). Discovering the Company of Surgeons. *The Bulletin of the Royal College of Surgeons of England*, *93*(6), 208–209. https://doi.org/10.1308/147363511x573923

Van De Ridder, J. M. M., Stokking, K. M., McGaghie, W. C., & Ten Cate, O. T. J.

    (2008). What is feedback in clinical education? *Medical Education*, *42*(2), 189–197.

    https://doi.org/10.1111/j.1365-2923.2007.02973.x

Van Der Horst, C. A., & Albertyn, R. M. (2018). The importance of metacognition and

    the experiential learning process within a cultural intelligence–based approach to

    cross-cultural coaching. *SA Journal of Human Resource Management*, *16*(1), 1–11.

Van Dongen, K. W., Ahlberg, G., Bonavina, L., Carter, F. J., Grantcharov, T. P.,

    Hyltander, A., … Broeders, I. A. M. J. (2011). European consensus on a

    competency-based virtual reality training program for basic endoscopic surgical

    psychomotor skills. *Surgical Endoscopy*, *25*(1), 166–171.

Vanderbilt, A. A., Grover, A. C., Pastis, N. J., Feldman, M., Granados, D. D. ia., Murithi,

    L. K., & Mainous, A. G. (2015, March 1). Randomized controlled trials: a

    systematic review of laparoscopic surgery and simulation-based training. *Global*

    Journal of Health Science, Vol. 7, pp. 310–327.

    https://doi.org/10.5539/gjhs.v7n2p310

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological*

    processes. Harvard university press.

Walker, M. P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J. A., & Stickgold, R.

    (2003). Sleep and the time course of motor skill learning. *Learning & Memory*,

    10(4), 275–284.

Wall, C. (1937). *The history of the Surgeons' Company, 1745-1800*. Hutchinson's

    scientific & technical publications.

Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: current state of

the art. *Advances in Health Sciences Education*, *7*(1), 63–80.

Warren, J. (1958). *To Work in the Vineyard of Surgery: The Reminiscences of J. Collins Warren, 1842-1927*. Harvard University Press.

Watson, D. R., Flesher, T. D., Ruiz, O., & Chung, J. S. (2010). Impact of the 80-hour workweek on surgical case exposure within a General Surgery residency program. *Journal of Surgical Education*, *67*(5), 283–289. https://doi.org/10.1016/j.jsurg.2010.07.012

Whitmore, J. (2010). *Coaching for performance: growing human potential and purpose: the principles and practice of coaching and leadership*. Hachette UK.

Wilasrusmee, C., Lertsithichai, P., & Kittur, D. S. (2007). Vascular Anastomosis Model: Relation Between Competency in a Laboratory-based Model and Surgical Competency. *European Journal of Vascular and Endovascular Surgery*, *34*(4), 405–410. https://doi.org/10.1016/j.ejvs.2007.05.015

Wilasrusmee, C., Phromsopha, N., Lertsitichai, P., & Kittur, D. S. (2007). A new vascular anastomosis model: relation between outcome and experience. *European Journal of Vascular and Endovascular Surgery*, *33*(2), 208–213.

Williams, W. H. (1972). The Early Days of Anglo-America's First Hospital: The Pennsylvania Hospital, 1751-1775. *JAMA: The Journal of the American Medical Association*, *220*(1), 115–119. https://doi.org/10.1001/jama.1972.03200010099019

Winfrey, M. L., & Weeks, D. L. (1993). Effects of Self-Modeling on Self-Efficacy and Balance Beam Performance. *Perceptual and Motor Skills*, *77*(3), 907–913. https://doi.org/10.2466/pms.1993.77.3.907