ON THE USE OF SIMULATION AND ASSESSMENT IN SURGICAL TRAINING

ON THE USE OF SIMULATION AND ASSESSMENT IN SURGICAL TRAINING

By NATALIE WAGNER, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2019) Hamilton Ontario

(Psychology)

TITLE: On the Use of Simulation and Assessment in Surgical Training

AUTHOR: Natalie Wagner, B.Sc.

SUPERVISOR: Dr. Ranil R. Sonnadara

PAGES: xx, 276

Lay Abstract

From day one, new doctors are expected to adapt to new training environments, manage increased workloads, and make decisions about patient care, while working longer hours, and with less support than they had as medical students. As such, research suggests that new doctors often feel underprepared and stressed when entering residency. Furthermore, the 'July Effect', which suggests that fatal medical errors rise by 10% in the first month of residency compared with all other months, suggests that something needs to change. In this thesis, we sought to improve the way that surgical training programs approach the transition into residency, and look at the long-term impacts of these changes. Through three interrelated studies, we provide educators with information on how to develop, implement, and evaluate a simulation-based boot camp for new doctors; highlight new approaches for evaluating educational initiatives; and revisit how assessment is being used in medical education.

Abstract

The transition from medical school to residency is often considered the most difficult year for both teachers and learners. Learners report feeling underprepared, and some researchers have identified a decrease in patient safety during the first month of residency. These factors suggest learners could be better supported during this transition period. Previous research demonstrates that boot camps (BCs) at the onset of residency can improve learners' confidence, knowledge, and some technical skills. However, little information has been published on how those BCs were developed and implemented, why BCs only improve some skills and not others, or the long-term impacts of BC programs.

We used a Context, Input, Process, and Product program evaluation framework to develop, implement, and evaluate a simulation-based BC for novice surgical trainees that was aligned with the recent shift towards competency-based models of medical education. Next, we used a Convergent Parallel Mixed Methods approach to explore the longer-term impacts of the BC program. Lastly, we explored how effectively the Objective Structured Clinical Examination (OSCE), a "gold standard" measure of learner competence that was used in the BC program, truly captures clinical performance of novice trainees.

This work demonstrates that incorporating a BC at the onset of residency can improve residents' confidence and skill for up to two years into training, although adherence to sound pedagogical principles is critical. The BC also provided residents with the opportunity to participate in role clarification, acculturation, and social integration. Finally, we demonstrate that OSCEs may not always be the best way to measure BC effectiveness.

The data presented in this thesis will provide educators with new insights on how to create and evaluate successful BC programs to support learners through the transition to residency; highlight new approaches for evaluating educational initiatives; and prompt a conversation about how assessment is being used in medical education.

Acknowledgements

It is difficult to put into words how grateful I am for the support I have received over the past five years, but there are a few people in particular I would like to thank.

To Vickie Galea, my first graduate supervisor, thank you for believing in me and welcoming me to McMaster. From our initial Skype conversation about the Infant Movement Project until now, your support and guidance have been unwavering. I am also extremely thankful that once I moved over to PNB, I was still able to call you a mentor. Your determination, ability to cope with the challenges of collecting data from newborns, and vivacious spirit will stay with me forever.

To Ranil Sonnadara, my graduate supervisor, thank you for introducing me to the world of medical education. Your high standards, attention to detail, and ability to generate multiple new research ideas within a single conversation has shaped me into the researcher I am today. I also want to thank you for your encouragement to pursue academic experiences outside of my thesis work. With your support (and your many hats) I have had the opportunity to be involved in so many different things; this allowed me to make the most out of my time at McMaster.

To my supervisory committee, past and present, thank you for your guidance with my thesis projects. In both cases, having clinician and researcher perspectives was invaluable.

To my fellow graduate students, I credit much of my success to you. Sarah and Sarah from my initial Rehabilitation Science cohort - although our academic paths went their separate ways, we have become lifelong friends. Our relationships, built on that first statistics course, have become so much more. To my PNB cohort, you welcomed me into a new research field and community with open arms. It has been a lot of fun to get to know everyone, win a volleyball championship, and struggle through those softball games. In particular, to Brendan and Hanae, despite our very different research interests, you made me feel like part of the group. I value your friendship, willingness to meet up for coffee (or beer), and discuss academics, or life. To the Sonnadara Lab – Anita, Portia, Tina, Sydney, Victoria, Nathan, Maverick, Sabahat, Jeni, and Jacqueline, there is so much I could say. From the early morning meetings at the hospitals, lab meetings, journal clubs, conferences, taco dates, and many Slack conversations, you made this a fun five years. Thank you for being an awesome team.

To everyone from the Department of Surgery at McMaster, thank you for believing in this research initiative. I will be forever indebted to those who took time away from their busy clinical schedules to assist with behind-the-scenes coordination, facilitate boot camp sessions, evaluate residents, and participate in interviews and focus groups. Included in that, I want to give a special thanks to Nalin Amin and Jane Klie, the Program Director and Program Coordinator of the Surgical Foundations program respectively - you both were part of this journey from the first boot camp planning meeting, and without your support none of this would have been possible.

To my family and friends, thank you is not enough. Despite not always knowing what program I am in, or exactly what I am studying, you have always believed in me. You have let me rant about statistics or referencing software without complaint, have organized celebrations around my academic schedule, Facetimed at any hour of the day, and have provided the life balance that I needed these past five years. I love you all.

Lastly, to Andrew, you have certainly received the brunt of my stress and ramblings these past five years. Thank you for being my best friend, my sounding board, my guinea pig, and my biggest support. Through the many degrees, houses, late nights at the library, adventures with Mya, drives to see family and friends, and recreational sports, I could not imagine doing this with anyone else by my side. Thank you for keeping everything in perspective.

Table of Contents

List of Tables, Figures, and Appendices

3.0 Context, Input, Process, and Product Evaluation Model to Develop, Implement, and Evaluate a Boot Camp for Novice Surgical Trainees

4.0 Long-Term Effectiveness of a Boot Camp for Novice Surgical Trainees: A Convergent Parallel Mixed Methods Study

5.0 The Objective Structured Clinical Exam as an Indicator of Clinical Performance

List of Abbreviations

ANOVA: Analysis of Variance

ATLS: Advanced Trauma and Life Support

BC: Boot Camp

BID Method: Briefing, intraoperative, debriefing method

CBME: Competency-based medical education

CBD: Competence By Design

CIPP: Context, Input, Process, and Product

CPMM: Convergent Parallel Mixed Methods

DOPS: Direct Observation of Procedural Skills

EPA: Entrustable professional activity

HiREB: Hamilton Integrated Research Ethics Board

ICC: Intraclass correlation coefficient

MCC: Medical Council of Canada

Mini-CEX: Mini-clinical evaluation exercise

MSF: Multi-source feedback

OB/GYN: Obstetrics and Gynecology

OSATS: Objective Structured Assessment of Technical Skills

OSCE: Objective Structured Clinical Exam

O-SCORE: Ottawa Surgical Competence Operating Room Evaluation

OTL-HNS: Otolaryngology, Head and Neck Surgery

PGY: Postgraduate year

QCA: Qualitative content analysis

QD: Qualitative description

RCPSC: Royal College of Physicians and Surgeons of Canada

SF: Surgical Foundations

TT: Traditionally-trained

WBA: Workplace-based assessment

Declaration of Academic Achievement

Natalie Wagner was responsible for all aspects of this thesis, including data collection, analysis, and thesis preparation.

Sydney McQueen and Ranil Sonnadara assisted with revisions of chapter two, which has been published as a book chapter: Wagner, N., McQueen, S., and Sonnadara, R.R. (2018). "Bridging the gap: theoretical principles behind surgical boot camps". R. Rambani (Ed.) *Boot Camp Approaches to Surgical Training.* (pp. 1-11). Basel, Switzerland: Springer.

Anita Acai, Christine Fahim, Jacqueline Wilcox, Jennifer Zering, Maverick Chan, Portia Kalun, Tianna Murray, and Tatiana Gayowsky assisted with data collection in chapter three.

Nathan Cupido assisted with data analysis in chapter four.

The thesis committee advised on all aspects of this thesis.

Chapter 1

General Introduction

1.0 General Introduction

1.1. Preface

> *A transition is not a moment, but rather a dynamic process where an individual moves from one set of circumstances to another… In coping with major changes, individuals face new challenges, opportunities, stress and a range of emotions. (Teunissen & Westerman, 2011, p. 52)*

Doctors experience many transitions during their training and ensuing careers (Kilminster, Zukas, Quinton, & Roberts, 2011; Teunissen & Westerman, 2011). Moving from the classroom to the clinical environment, rotating to a different specialty, changing to a new clinic or hospital, and moving between junior to senior learner roles all involve rapidly adapting to new contexts and situations, needing to acquire new knowledge, learning where to turn for help, and doing so while providing safe care to patients (Kilminster et al., 2011). Of the many transitions, the progression from medical school to residency is often considered the most challenging year for both teachers and learners (Cave, Woolf, Jones, & Dacre, 2009; Kilminster et al., 2011; Teagle, George, Gainsborough, Haq, & Okorie, 2017; Teo, Harleman, O'Sullivan P, & Maa, 2011). Previous research suggests medical trainees feel underprepared and stressed as they take on new roles and responsibilities at the start of residency (Bligh, 2002; A. Cameron, Millar, Szmidt, Hanlon, & Cleland, 2014; Cave et al., 2009). Furthermore, several studies have suggested that teaching hospitals experience increased mortality and decreased efficiency during doctors' first month of residency when compared with other months (Young et al., 2011). Taken together, there is a need to support learners more effectively as they transition into residency.

At the same time, there have been a number of changes to healthcare systems around the world. Recently, these changes have forced educators to re-evaluate how we are teaching medical trainees. This re-evaluation has resulted in a shift away from traditional models of medical education towards competency-based medical education frameworks. While this shift is primarily a response to the evolving needs of healthcare systems, it also brings forth an opportunity to explore ways to better support medical trainees during critical transition points, such as the transition from medical school to residency.

This chapter will (1) provide a brief overview of the history of medical education; (2) review the traditional, time-based model of medical training; (3) discuss changing demands on healthcare systems; (4) highlight challenges seen during the transition from medical school to residency; (5) introduce the idea of 'boot camps' for new medical trainees; (6) describe the shift towards competency-based medical education and the resultant opportunity for curricular change; (7) identify the gaps in the literature; and (8) define the goals of the present thesis.

1.2 History of Medical Education

Over the centuries, medical training has experienced many reforms. The history of these reforms is essential to understanding current issues in medical education and the context in which they developed.

1.2.1 Medical Education in Europe

In the Middle Ages, the sick and poor were typically cared for by those in the holy order, such as priests or monks (Robinson, 1984). However, in 1163 the Pope declared that blood shedding was an ‘unholy’ matter (Dobson, 1974; Robinson, 1984). As a result, medicine and surgery separated into two distinct fields. Medicine remained a respected profession, with medical training occurring in monasteries and academic institutes (Robinson, 1984). In contrast, surgery was deemed a trade and surgical procedures were delegated to barbers (Franzese & Stringer, 2007). Barbers were selected as they were both trusted to hold a razor to the neck of a royal person without inflicting harm, and already had most of the tools required for surgical procedures (Dobson & Walker, 1979). Surgical training took place through informal apprenticeships rather than formal, classroom-based teaching (Dobson & Walker, 1979).

It was not until the 1500s (over three hundred years later) that this system of medical training changed. While patients were visiting the barbers, it was common for them to receive treatment for other medical ailments. However, since barbers did not have academic training in medicine, many physicians were concerned that patients were receiving inadequate care. In an effort to combat this problem and create more defined borders between physicians, surgeons, and apothecaries (who were also known to provide treatments), the College of Physicians was formed (Dobson & Walker, 1979). In 1518 the College of Physicians (now known as the Royal College of Physicians) mandated that all physicians hold a University degree. From that point forward, the College of Physicians oversaw all medical training and licensure for physicians in London (Dobson & Walker, 1979).

Around the same time as the formation of the College of Physicians, the Guild of Barbers, the group of barbers who had taken over surgical practices in London, came in conflict with a group called the Fellowship of Surgeons. The Fellowship of Surgeons was a small, more select group of individuals who had received extensive training from master surgeons in other parts of Europe (Dobson & Walker, 1979). The issue was that both the Guild of Barbers and the Fellowship of Surgeons were claiming the right to regulate the practice of surgery in London (Dobson & Walker, 1979). While the Fellowship of Surgeons had more expertise, the barbers had the numbers, and had support from the church and royal family. Moreover, the Guild of Barbers had just lost their ability to treat medical ailments due to the formation of the College of Physicians; thus, they did not want to give up surgery that easily. After years of negotiations, in 1540 the two groups formed the Barbers' Company (Dobson & Walker, 1979). With the formation of the Barbers' Company, new rules were created. "No surgeon was to practice barbery and no barber could practice surgery, except the drawing of teeth" (Dobson & Walker, 1979, p. 34). All surgeons had to put a sign on the street so people knew where to go for help. Finally, surgical training and practice had to be monitored. The leaders of the Barbers' Company became responsible for approving apprenticeship positions, organizing anatomy lectures, granting licenses, and monitoring those practicing surgery (Dobson & Walker, 1979).

For over two hundred years, physicians from the College of Physicians, and barbers and surgeons from the Barbers' Company worked in harmony. However, in the early 1700s things once again changed. Surgeons wanted to regain the reputation of being an academic discipline like medicine (Dobson & Walker, 1979). In hopes of accomplishing this goal, the surgeons separated from the Barbers' Company in 1745 and created their own College of Surgeons (now known as the Royal College of Surgeons) (Dobson & Walker, 1979). Following the separation, barbers maintained their teaching halls and the traditions of the company (to this day the Barbers' Company functions as a charitable organization), but the surgeons finally gained claim over surgical practice. From that point onwards, surgeons worked towards restoring surgery to a respected profession in Europe (Dobson, 1974).

Understanding the fact that surgery and medicine were separate fields for several centuries is critical to understanding why some aspects of medicine are the way they are today, and is something that will be revisited later in the thesis. However, for the purpose of this thesis, we will shift our focus from medical education in Europe, to explore how medical education developed in North America.

1.2.2 Medical Education in America

During the colonization of the Americas in the 1600s and 1700s, many physicians and surgeons moved their practices from Europe to America. However, there were no regulatory bodies to oversee medical training in America as there were in Europe. Not having a governing body meant that new physicians and surgeons were once again trained through informal apprenticeships with no rules or regulations (Flexner, 1910). Some individuals travelled back to Europe after their apprenticeship for additional training; however, not many could afford to make the trip (Flexner, 1910). In 1765 John Morgan suggested that a medical school be created in America for those that could not afford to travel to Europe and back (Flexner, 1910). Morgan proposed that medical schools could supplement apprenticeship training by providing a venue for those who had travelled to

Europe to disseminate their knowledge to American trainees (Flexner, 1910). James Bond, a founder of the Pennsylvania Hospital, built on this idea and suggested that as a part of medical school, trainees should have the opportunity to learn from real patients in the hospital environment (Flexner, 1910). Their proposals were met with support, resulting in the formation of the first medical school and teaching hospital in America (Flexner, 1910). Between 1765 and 1800, three other medical schools were created that followed this format. However, due to the American Revolution, further developments in medical education were temporarily put on hold (Flexner, 1910).

Following the American Revolution, medical schools began to pop up all over the country. Unfortunately, as the number of medical schools increased, the structure of the original four medical schools was lost. Anytime there were more than six physicians in one city, they were likely to create a medical school. In turn, new medical schools were often not associated with a University or College, included no clinical components, and varied in length; while apprenticeships had disappeared entirely. Medical school had turned into a business venture - anyone who could pay to attend was accepted. In the late 1880s, there was growing concern that the lack of standards meant medical schools were producing too many undertrained physicians and surgeons. In response to these concerns, the trustees of the Carnegie Foundation, a group established by Congress to uphold higher education, commissioned a study on medical training in 1908. The investigators on this study visited every medical school in the United States and Canada to describe the status of training. The main investigator on the study, Abraham Flexner, detailed his findings in a report in 1910. This report changed medical education forever.

Flexner described a general lack of preparedness amongst medical graduates, a lack of standards across medical schools, and a lack of funding and support from Universities and Colleges. Based on these findings, Flexner provided recommendations for improving the medical education system. These recommendations included the incorporation of basic sciences and clinical training in medical school; standardizing the medical school curriculum; and having Universities and Colleges responsible for upholding these standards and supporting medical schools. Flexner emphasized that these recommendations were minimums, not a visionary ideal. He also emphasized that how physicians and surgeons are trained should be of great importance not only to medical practitioners, but to all citizens whose well-being relies on the quality of training (Flexner, 1910). With the exception of a standardized medical school curriculum, all of Flexner's recommendations were realized in the years to follow. This was, in part, due to the support gathered from the 'Flexner Report', but also as a result of other prominent individuals in medical education at the time - Sir William Osler and William Halsted.

Prior to Flexner's report, Sir William Osler, the Chief Physician at Johns Hopkins Hospital, proposed that one way to address the concern that physicians were undertrained, was to link the academic study of medicine to functioning hospitals (Roland, 1982). Osler suggested that the first two years of medical school should teach the basic sciences in classroom and laboratories, while the last two years of medical school should have students apply their knowledge in clinical placements (Roland, 1982). Osler's idea was reminiscent of the format in the original four medical schools and, combined with Flexner's recommendations for structured medical school, formed clinical clerkships.

In addition to the formation of clinical clerkships, William Halsted, Osler's colleague and the Chief Surgeon at Johns Hopkins Hospital at the time, introduced a system where medical school graduates could apply for additional, hospital-based training in surgery (J. L. Cameron, 1996). Halsted proposed that over a period of seven years, trainees could be granted increased responsibility for treating patients (J. L. Cameron, 1996). Halsted's program retained an apprenticeship style, similar to what was being done in Europe and before the commercialization of medical schools in America. The difference was that Halsted's program occurred *after* medical school. Through this program Halsted aimed to raise the standard of surgery, which still was not considered an academic discipline in the way that medicine was viewed (J. L. Cameron, 1996). Halsted's work provided additional momentum for Flexner's recommendations, and was the beginning of postgraduate training (J. L. Cameron, 1996).

1.3 Traditional (Time-Based) Model of Medical Education

Since these contributions by Flexner, Osler, and Halsted over one hundred years ago, there have been few significant changes to medical training (Hodges, 2010). In today's training model, all aspiring physicians and surgeons must first complete medical school. Medical school ranges from three to six years (depending on the country) (Teunissen & Westerman, 2011). Medical school curricula involve a period of pre-clinical teaching, typically in a classroom setting, followed by a period of clinical clerkship that spans all major domains of medicine (Teunissen & Westerman, 2011). After completing medical school, postgraduate training (sometimes called 'residency training') is mandatory, where learners spend a set period of time working in the clinical environment under the supervision of 'expert' clinicians or surgeons in a focused area or 'specialty' (R. Sonnadara et al., 2013). While Halsted started postgraduate training for surgery specifically, postgraduate training is now required for those wishing to pursue a career in medicine or surgery, although the length of postgraduate training varies by specialty. Once trainees have completed the predetermined training period, they are eligible to complete examinations which, if successful, make them qualified to hold a license for independent practice in their area of specialization (Association of American Medical Colleges, n.d.).

Notably, each phase of this traditional model of medical education is *time-based.* Hodges (2010) likens the traditional model of medical education to 'tea steeping'. Some things take time to learn and cannot be taught in the classroom. Thus, it is only after having been immersed in the clinical environment for a certain amount of time that trainees will have gained enough exposure and experience to be prepared to move to the next step of medical training, and ultimately for independent practice in their given area. This time-based, or 'tea-steeping', model has been extremely durable, evidenced by the few changes to medical education over the past 100 years (Hodges, 2010). However, as demands on healthcare systems are changing, it seems another wave of reform might be upon us.

1.4 Need for Change

Due to increases in life expectancy and the 'baby boom', the aging population is growing (British Medical Association, 2018). This growth means that healthcare systems are caring for more people and care needs are becoming more complex as individuals live longer (British Medical Association, 2018). Due (in part) to these complex care needs, scientific discoveries and technological advancements in healthcare are ongoing. As a result, the knowledge required of healthcare workers continues to increase (Gawande, 2010). To quote physician Atul Gawande, "Clinicians now have at their disposal some six thousand drugs and four thousand medical and surgical procedures, each with different requirements, risks, and considerations. It is a lot to get right" (Gawande, 2010, p. 35). As Gawande published his book in 2010, one can only imagine how these numbers have risen.

Furthermore, with the increasing patient population there are mounting pressures from hospital administrators, society, and the doctors themselves, to get people treated as fast as possible (British Medical Association, 2018). Due to this push for efficiency, it is becoming increasingly difficult to carve out meaningful teaching time for medical trainees (Qureshi & Maxwell, 2012). With the increased workload, doctors are also exhibiting higher stress levels and are at an increased risk of burnout (British Medical Association, 2018). Burnout, as described by Shanafelt et al. (2015, p. 1601), is "a syndrome of emotional exhaustion, loss of meaning in work, feelings of ineffectiveness, and a tendency to view people as objects rather than human beings". Recent literature suggests that over 50% of physicians experience burnout (Shanafelt et al., 2015), and that this burnout can undermine professional development, place patients at risk, and lead to personal consequences such as substance abuse and suicide ideation (Dyrbye & Shanafelt, 2016). In fact, prevalence of suicide among physicians is estimated to be higher than that among the population in the United States, despite similar levels of depression (Dyrbye &

Shanafelt, 2016). Burnout is a concern for all physicians, but particularly for junior residents who perform many overnight call shifts, during which there are numerous patient demands, a decreased number of staff, and expectations to complete work independently despite little clinical experience (British Medical Association, 2018; Gulland, 2017). With junior residents having so much responsibility while feeling both physically and emotionally drained, there are concerns that patient safety may be negatively impacted (Gulland, 2017).

One such example is the Libby Zion case from 1984 (Block, 1994; Lerner, 2009; N. Patel, 2014; Sullivan, 1987; Swadron, 2018). Libby Zion was an 18-year-old college student in New York that had been sick for several days. Her family doctor recommended she go to the emergency department, and when she arrived at 11:30 p.m. that night, Libby had a high fever and was not making sense. At 2:00 a.m. a second-year resident admitted her to the internal medicine floor for observation. At 3:00 a.m. Libby was assessed by a first-year resident and the same second-year resident as earlier, diagnosed with a viral infection, and given a painkiller to ease her discomfort. The staff physician was consulted over the phone and approved of the plan. The second-year resident left the hospital to get some sleep (as was customary), and the first-year resident took over responsibility of the patients on the floor. At 3:30 a.m. the first-year resident received a call from the nurses saying Libby had become agitated. Concerned that Libby may hurt herself, the first-year resident ordered restraints, additional painkillers and an antipsychotic medication. When Libby's vital signs were checked again at 6:00 a.m., she had a fever of 107F. The nurses immediately made efforts to cool her body, but Libby went into cardiac arrest and could not be revived. Within the span of seven hours, Libby Zion had passed away.

The original cause of death was stated to be pneumonia. However, when Libby Zion's father found out that the staff physician never saw her, the first-year resident did not re-evaluate her during the night, and residents often worked 36 hour shifts with no sleep, he called for an investigation (Block, 1994; N. Patel, 2014). After re-examination, the cause of death was suspected to be a fatal drug interaction (Block, 1994; Swadron, 2018). As a result, Libby Zion's story was put forward to a grand jury as a malpractice case (Sullivan, 1987). In a somewhat surprising verdict, the grand jury refused to indict the physicians involved, but assigned equal blame to the hospital system and to Libby Zion (Sullivan, 1987). Libby should have been taken to the intensive care unit and undergone different diagnostic tests. However, Libby may have concealed illicit drug use that contributed to her symptoms, and the residents were working long hours with little supervision or support (Sullivan, 1987). The jury's main concern was that this working environment was not case-specific, rather it existed for residents in all teaching hospitals (Sullivan, 1987). The grand jury called upon the state department of health to issue regulations to ensure greater supervision of medical trainees (Sullivan, 1987). Soon after, the state department issued a number of recommendations including, staffing emergency departments with licensed physicians who have at least three years of clinical experience; having an attending physician supervise all junior residents; and placing regulations on the number of hours learners can work in teaching hospitals (N. Patel, 2014).

While not all of these recommendations were put into practice, this case propelled several countries to move forward with resident work-hour limitations. For example, trainees in Canada are now restricted to 60-90 hours/week depending on the province (National Steering Committee on Resident Duty Hours, 2013; Pattani, Wu, & Dhalla, 2014), trainees in the United States are restricted to 80 hours/week (Riebschleger & Philbert, 2011), trainees in New Zealand to 72 hours/week (New Zealand Resident Doctors' Association, 2007), and trainees in Europe to 48 hours/week (National Health Service Employers, 2009). Regulations on the length of call shifts, days off per week, and amount of time between shifts have also been made (N. Patel, 2014).

Together, these changes to the healthcare system, including the increased number of patients, push for increased efficiency, decreased time for teaching in the clinical environment, increased knowledge required of trainees, and decreased number of resident work hours, represent a number of competing demands. As such, educators are concerned that the traditional, time-based model of training can longer ensure that 'x' amount of time is sufficient for trainees to learn everything they need to know for independent practice (R. Sonnadara et al., 2013). Highlighting this issue, Bell et al. (2009) compared a list of procedures that general surgery Program Directors in the United States considered to be essential, with the actual experience of graduating trainees. Their results suggested that of the 121 procedures deemed essential, the average graduating trainee completed one of the procedures more than 50 times (laparoscopic cholecystectomy) and 83 of them less than five times. Additionally, there were 63 procedures where most trainees reported no experience at all. This study brought attention to two things: first, operative experience is not standardized nationally; and second, surgical trainees do not have adequate operative experience prior to graduation (Bell et al., 2010).

One simple solution might be to make training longer. However, in North America physicians are already completing four years in an undergraduate degree, four years in medical school, and five years in residency. This training structure means that on average, physicians are not starting to work until they are at least 31 years old. This estimate does not include time off between degrees, additional graduate degrees, time spent sub-specializing after residency (fellowship), or parental leave. With some researchers suggesting cognitive decline is present by 45 years old (Singh-Manoux et al., 2012), lengthening training would mean trainees are not starting to work until after their prime. Additionally, the financial costs associated with medical education are already extremely high (Dorsey, Nincic, & Schwartz, 2006); lengthening training would only propagate the financial burden. Lastly, even if training was lengthened, the results from the Bell et al. (2010) study suggest we cannot assume that 'x' amount of years in training results in competent trainees. Within the same amount of time spent in training, trainees can have dramatically different clinical experiences. Something else has to change - we need to find ways to make the current system more effective. In looking for solutions, we focus in on some of the critical gaps in medical education.

1.5 Transitions in Medical Education

Many critical gaps appear around transition points in medical education. In their review paper, Teunissen and Westerman (2011) highlight some of the challenges during three main transition points in medical education. When medical students begin their clinical part of training (clerkship) they often have difficulty applying their theoretical knowledge to practice and struggle to understand their role and responsibilities in the clinical environment. When medical students begin residency, there is a sudden increase in workload and decrease in studying time. Furthermore, one in four new residents develop burnout as a result of the responsibility associated with patient care. When trainees become independent practitioners, they must learn how to cope with the financial aspects of patient care and responsibility for training other doctors. These three transition points are common to all specialties; however, for those in longer programs, there is an increasing awareness of another transition occurring midway through residency (Strohnbehn, Cronin, & Levy, 2018). For programs that are five years or more, trainees are typically considered a junior resident in postgraduate years (PGY) 1-2 and then a senior resident in PGY 3 -5. When trainees take on the role of senior resident, there is the expectation that they will be more autonomous, support junior residents, and be able to complete more complex clinical skills (Bohnen, 2018).

Each of these transitions comes with its own challenges and highlights an area where training could be more effective. However, since the transition from medical school to residency is often reported as the most difficult year for both teachers and learners (Bligh, 2002), we will focus the remainder of the thesis on that transition point.

1.5.1 Transition From Medical School to Residency

Starting on day one of residency there is an expectation that new doctors will perform adequately in their new clinical role, make decisions regarding patient care, and manage their new responsibilities (Prince, Van de Wiel, Van der Vleuten, Boshuizen, &

Scherpbier, 2004). Research suggests this transition often makes trainees feel underprepared and stressed when entering residency (Bligh, 2002; Kilminster et al., 2011; Prince et al., 2004; Teo et al., 2011). The combination of being underprepared and stressed not only places trainees in a vulnerable position, but also poses a risk to patient safety (Cohen et al., 2013). Because of this perceived risk to patient safety, there has been a lot of interest in whether the annual influx of new residents affects patient outcomes (Young et al., 2011).

1.5.2 The 'July Effect' and 'Black Wednesday'

In 2010, Phillips and Barker conducted a retrospective study of death certificates in the United States. Through this study, Phillips and Barker (2010) identified a decrease in health care efficiency and patient safety, with fatal medical errors rising by 10% in July compared to all other months. As this increase in fatal errors was only seen in counties with teaching hospitals, Phillips and Barker concluded that the 'July Effect' was due (at least in part) to changes associated with the arrival of new medical residents. Following this study, Young et al. (2011) conducted a systematic review of all studies examining the effect of medical trainee changeover on patient outcomes. The results from their systematic review agreed with Phillips and Barker, suggesting mortality increased and efficiency decreased in hospitals during the year-end changeover in medical trainees (Young et al., 2011).

Similar studies have also taken place in the United Kingdom, where the academic year for medical trainees starts on the first Wednesday in August rather than July 1st. For example, Jen, Bottle, Majeed, Bell, and Aylin (2009) conducted a retrospective study on patients admitted to hospitals in England the last week of July versus the first week of August from 2000 to 2008. Their results suggest that those admitted in August had a 6-8% increased mortality rate compared with those admitted in July when controlling for year, gender, age, socio-economic status and other comorbidities (Jen et al., 2009). A report by Dr. Foster Intelligence in the United Kingdom provided further support for 'Black Wednesday', as they evaluated 300,000 emergency admissions over nine years and once again found a 6-8% increase in mortality in the first week of August compared to the last week of July (Gaskell, Hinton, Page, Elvins, & Malin, 2016; Imperial College of London, 2009). While it is challenging to determine the factors that make the trainee changeover time so problematic (Young et al., 2011), these findings suggest among all the changes to the healthcare system, the transition from medical school to residency is one of the critical points in which learners need to be better supported.

1.6 Boot Camps in Medical Training

Some institutions have implemented an intensive skills course, or "boot camp" at the onset of residency to support learners through the transition from medical school to residency (Blackmore, Austin, Lopushinsky, & Donnon, 2014). While different formats exist, most boot camps (BCs) take place across several consecutive days and include a combination of didactic lectures, small-group activities or tutorials, lab-based technical skills sessions, and simulations (Safir, Sonnadara, Mironova, & Rambani, 2018). Research suggests this type of training program can be highly effective, as residents who participate in a BC can score significantly higher on measures of surgical skill compared with those who embarked on traditional residency (Cohen et al., 2013; R.R Sonnadara et al., 2011). Boot camps can allow junior residents to achieve the same technical skill level as senior residents in a laboratory setting (R. R. Sonnadara et al., 2012). Lastly, residents can maintain the skills acquired in a BC for up to seven months, suggesting that skills learned through a BC have an excellent retention rate (R. R. Sonnadara et al., 2012). Blackmore et al. (2014) conducted a meta-analysis to assess the effectiveness of BCs during the transition to residency and reported that learners who completed BCs had significant improvements in clinical skills, knowledge, and confidence.

Despite these promising results, BC research is still in its infancy and not all programs have implemented BCs (Blackmore et al., 2014). This may be due to a variety of barriers, such as: requiring someone with the time and expertise to organize and implement the BC (Safir et al., 2018); ensuring adequate patient care when trainees and staff physicians are absent from clinical service to attend BCs (Fernandez et al., 2012; Safir et al., 2018); and lastly, cost. Satterthwaite, Leighton, Safir, and editors (2018) estimate that the University of Toronto Surgical Prep Camp program costs at least CDN$1000 for each trainee. As the University of Toronto has over 50 incoming trainees each year, running a boot camp can incur significant cost. For programs with fewer residents, boot camps may not be as costly; however, the program may not have the manpower to simultaneously run a BC and cover clinical service. Thus, strong buy-in from hospital administrators and educators is required to overcome these barriers and implement a BC. Historically, that buy-in has not been present. However, with the recent changes to healthcare systems (as discussed in 1.4), there is need for another drastic reform to medical education. With this reform, there may be an opportunity to implement changes and address transitions in a way that was not previously possible.

1.7 Shift to Competency-Based Medical Education

1.7.1 Development of the CanMEDS Roles

In 1990, the Royal College of Physicians and Surgeons of Canada (RCPSC) commissioned a project to help medical education meet the changing needs of society (Frank, Snell, Sherbino, & editors, 2015). The result of this project was the development of CanMEDS, a framework outlining seven key roles of an effective physician (medical expert, manager, health advocate, collaborator, communicator, scholar and professional). Following its launch in 1996, CanMEDS became popular with healthcare programs around the world (Frank, Snell, & Sherbino, 2015). However, with growing concerns that the traditional, time-based model of medical education could no longer ensure trainees gain adequate clinical exposure prior to graduation (as discussed in 1.4), the CanMEDS framework was revised in 2015 and evolved into a competency-based medical education framework.

1.7.2 Competency-based Medical Education

A true competency-based medical education (CBME) framework allows trainees to progress through their training based on the demonstration of competence, rather than fulfillment of a predefined amount of time (O. ten Cate, 2017). However, it would be nearly impossible to provide patient care if trainees moved on to a new rotation as soon as they demonstrated competence. Imagine arriving at the emergency department, only to find no trainees present since they already passed their rotation. To avoid this issue, most countries are proposing a hybrid model of CBME in which trainees still spend a predetermined amount of time on each rotation. However, within that time period (e.g. 4 weeks), trainees are responsible for achieving certain 'competencies'. Those who achieve the competencies may progress onto more advanced skills within that time, while trainees who experience difficulty will be given the opportunity to receive extra support where needed (Frank et al., 2010). This hybrid version of CBME has the potential to individualize learning and focus residency training on the learner, while still meeting service provision needs.

While frameworks exist under different names, for example Competence By Design (CBD) in Canada (Frank, Snell, & Sherbino, 2015), the Outcomes Project in the United States (Accreditation Council for Graduate Medical Education, 1999), and Common Competencies Framework for Doctors in the United Kingdom (Academy of Medical Royal Colleges, 2009), the hybrid version of CBME follows similar tenets across countries. For the purpose of this thesis we will focus on the Canadian context.

1.7.3 Competence By Design

When the CanMEDS roles were being revised in 2015, the Royal College of Physicians and Surgeons of Canada (RCPSC) announced that all postgraduate medical training programs in Canada would be required to implement the Canadian CBME framework, entitled Competence By Design (CBD) (Frank, Snell, Sherbino, et al., 2015). To accomplish this, the RCPSC proposed a phased rollout schedule, where the different medical and surgical specialties were to gradually adopt CBD over five years (Royal College of Physicians and Surgeons of Canada, 2018).

To spearhead the transition, the RCPSC proposed a national CBD committee be created for each specialty (Royal College of Physicians and Surgeons of Canada, 2019). The CBD Specialty Committees were to involve all Program Directors for that specialty, clinician educators, continuing professional development experts, and invited guests. The responsibility of the CBD Specialty Committee would be to identify the tasks in which trainees must be competent prior to graduation (centered around the 2015 CanMEDS roles), define how those tasks would be assessed, and revise the national standards in accordance. The RCPSC proposed that the CBD Specialty Committees would create these guidelines at the national level; however, how CBD was actually implemented in each residency program could vary according to local contexts.

1.7.4 Entrustable Professional Activities and Milestones

In CBD, the tasks that trainees must be deemed competent on are called Entrustable Professional Activities (EPAs). Entrustable Professional Activities are tasks that a clinician does as part of their daily work (O. ten Cate, 2018). Each EPA consists of several components or "milestones" that should be accomplished throughout training. For example, a trainee must be able to obtain informed consent from a patient (this would be an EPA). Obtaining an informed consent may be broken down into several components; for example: reviewing the risks, the benefits, and confirming the person understood (these would be the milestones). Some milestones may be accomplished quickly, some may take longer, and some may be context-dependent (e.g. obtaining consent for a simple versus complex procedure, or obtaining consent from a power of attorney). Once competence has been demonstrated on all milestones and in the various contexts, a trainee may be entrusted to complete that task independently. Once trainees have demonstrated competence on all of the EPAs pre-determined by their CBD Specialty Committee, they may be entrusted for independent practice and are eligible to graduate.

1.7.5 Assessment as the Cornerstone of CBD

If trainees are to be entrusted for independent practice based on the demonstration of competence, a fundamental requirement of CBD programs is the ability to measure 'competence'. In order to do that, there is a need for comprehensive assessment practices. Historically, assessment has been split into two types: formative and summative. Assessment described as "formative" is low-stakes, and takes place on an ongoing basis, guiding learners through immediate feedback (Sadler, 1989). In contrast, assessment described as "summative" is typically infrequent, high-stakes, and serves to determine if something has been achieved or not (Ferris & Flynn, 2015).

Prior to CBD, most medical training programs focused on summative assessments. Residency programs did these assessments through end-of-rotation evaluations that, in many cases, were accompanied by little to no feedback and were often completed well after the rotation had ended (R. Patel, Drover, & Chafe, 2015). Not surprisingly, there have been many documented issues with this summative assessment approach in medical education. Previous research has found that supervising physicians and surgeons are often not willing to assign low marks to poorly-performing trainees on summative assessments, creating a 'failure to fail' phenomenon (Dudek, Marks, & Regehr, 2005). This failure to fail occurs for a number of reasons, including: insufficient documentation of trainee performance; a lack of support from the program; insufficient observation by faculty;

competing time demands; fear of appeals and legal action; and fear of reciprocated poor faculty evaluations (McQueen et al., 2016). If assessment is to be the cornerstone of CBD, it is clear this summative approach will not work and something has to change.

In light of this concern, educators have proposed that formative assessment should be the primary driver of curriculum and learning within competency-based frameworks (Holmboe, Sherbino, Long, Swing, & Frank, 2010). Taking a formative approach would mean that individual assessments provide feedback to learners in a timely manner, and any one assessment does not weigh on their progression. Rather, when all of the assessments are brought together in a portfolio (or program of assessment), the data points can be aggregated to provide educators with a sense of the trainees' competence. As such, CBD requires a number of formative assessments, completed in the clinical environment or simulation, for each EPA. After seeing the aggregated EPA data, educators make a summative decision regarding trainee progression (van der Vleuten et al., 2012). The hope is that this approach will mitigate the failure to fail phenomenon, result in a more accurate representation of trainee skill, and ensure residents who may need additional support are identified early on in training.

1.7.6 Stages of Training

Recognizing that medical trainees are given increasing responsibilities based on previous experience and competence, and that this is only partially related to the amount of time spent in training, CBD has broken down postgraduate training into four new stages: transition to discipline, foundations of discipline, core of discipline, and transition to practice (Royal College of Physicians and Surgeons of Canada, 2016). Changing the terminology from postgraduate year (PGY) to stage of training is meant to help shift mentality from a time-based system, to a competency-based system. However, this new terminology also highlights the importance of transitions. As discussed in section 1.4, we are increasingly aware of how critical transition periods are in medical education. Having the CBD curriculum focus on the transition from medical school to residency through the transition to discipline stage; the separate roles of junior and senior residents through the foundation and core stages; and the transition from resident to independent practitioner through the transition to practice phase, provides an opportunity to focus curricular changes on these transition periods.

1.7.7 Opportunities for Curricular Change

One program where curricular change might be particularly effective is Surgical Foundations (SF). Surgical Foundations (SF) is part of the Canadian curriculum for all first- and second-year surgical residents. This includes those in Cardiac Surgery, Plastic Surgery, Obstetrics and Gynecology (OB/GYN), Vascular Surgery, Urology, Otolaryngology – Head and Neck Surgery (OTL-HNS), Neurosurgery, General Surgery, and Orthopaedics (Royal College of Physicians and Surgeons of Canada, 2010). The goal of SF is to provide trainees with the basic information and skills required for further training in a surgical specialty or subspecialty. Surgical Foundations (SF) programs typically hold educational sessions, referred to as academic half-days, once per week. Outside the SF academic half-days, residents are responsible for working and learning in the clinical environment at the direction of their home program (McMaster University, 2017). The SF program is in a unique position to implement curricular change for a few reasons.

The SF curriculum targets learners at the beginning of residency. Thus, SF is in a prime position to support learners through the transition from medical school to residency, which we know is one of the most challenging transitions in medical education. Second, SF is the only educational program where residents from different specialties come together to learn the same material. Recall from 1.2, it was only recently that surgeons started attending medical school with physicians. For that reason, medical school is still largely focused on teaching general medicine; apart from some clerkship rotations, medical students get limited exposure or teaching on surgery. Thus, surgical training programs must teach learners the basic principles of surgery once they enter residency. The SF program was created so that this teaching can be provided to trainees from all surgical specialties at once. As such, a large number of learners can benefit from any curricular changes to SF. Moreover, the SF program can collaborate with faculty and staff from the different surgical specialties, increasing opportunities for faculty development, buy-in, and manpower in general. Lastly, as SF was scheduled to be one of the first programs to transition to CBD in Canada, there has been urgency amongst program directors and administrators for change.

Taken together, the SF program has many qualities that place it in a unique position to implement curricular change. Implementing a BC type program in SF could target the transition from medical school to residency and support a large number of learners from different specialties. Additionally, due to the transition to CBD, there is now buy-in to pool resources and make these changes. However, before BCs can be implemented, gaps in the BC literature must be addressed.

1.8 Gaps in the Literature

As previously mentioned, research on BCs is still in its infancy. To date, research on the efficacy of BC programs has almost exclusively focused on quantitative measures of performance over a short time period. Many studies use pretest/posttest designs to explore whether the BC was effective (Blackmore, Puligandla, Emil, Romao, & Lopushinsky, 2019; Fernandez et al., 2012; Peyre, Peyre, Sullivan, & Towfigh, 2006). However, very few of those include a control group (Blackmore et al., 2014). In those that do use a comparator group, results suggest that BCs can improve *some skills* over a short period of time, but those skills tend to vary between studies, and many studies omit critical information on why that may be the case. For example, Parent et al. (2010) compared performance between residents who completed a BC and those who completed traditional training, on four surgical tasks (chest tube insertion, central line placement, wound closure, and a Fundamentals of Laparoscopic Surgery peg transfer task) at baseline, one month, and six months into training. Their results suggest that the BC cohort performed better on some of the tasks (e.g. chest tube insertion and central line placement) but not others (e.g. wound closure and the peg transfer task). Similarly, R. R. Sonnadara et al. (2012) measured effectiveness through comparing performance of the BC cohort with performance of other junior residents who completed traditional training on nine different surgical tasks immediately following the BC, and seven months later. Results from this study indicate that the BC cohort performed significantly better in all tasks except prepping and draping. These findings leave us with the question of why differences are seen between BC groups and comparator groups for some tasks but not others.

A few studies have tried to determine why this happens through collecting qualitative data on learners' experiences during a BC. For example, Chin, Roth, Rotenberg, and Fung (2014) conducted a pre- and post-BC survey and a structured follow-up telephone interview to evaluate residents' experience in the BC. Similarly, Cleland, Walker, Gale, and Nicol (2016) conducted an ethnographic study to explore the socio-cultural complexity of surgical BCs. These studies identify a variety of non-skill-based factors that influence the 'success' of a BC, such as having opportunities to meet their peers, gain 'insider information' from senior residents on how to excel in residency, meet faculty members in a less stressful environment, and learn some of the hospital logistics with which they may not be familiar (Cleland et al., 2016). However, they have provided little information on why certain skills are effectively taught through BCs, while others are not.

One way of addressing this question may be through mixed methods research. Mixed methods research is a way to collect, analyze, and integrate both qualitative and quantitative data in a purposeful manner (Creswell & Plano Clark, 2018). Creswell (2007) suggests mixed methods research provides a way to emphasize the strengths of quantitative and qualitative methods, while also offsetting their weaknesses. For example, quantitative methods are typically objective and generalizable; yet lack information on the context. Qualitative methods highlight the contextual piece, however are criticized for the subjectivity and lack of generalizability. Through combining approaches, we are able to answer questions that cannot be answered by one method alone. For example, through combining the qualitative data on BC effectiveness with quantitative measures of performance, educators may be able to capture what skills BCs are effective in teaching, and what contextual factors contribute to that effectiveness. To date, this mixed method approach has not been adopted in the BC literature.

In addition to this method issue, there has been little information on how BC programs were developed and implemented (Blackmore et al., 2014). The lack of information limits our understanding of why certain aspects would have been effective or not and the generalizability of the findings, as others are unable to recreate the BC within their own training programs. Lastly, as the longest follow-up period in these studies has been seven months (R. R. Sonnadara et al., 2012), there is a need for studies to explore the long-term impact of BC programs (Blackmore et al., 2014). This brings us to the goals of the present thesis.

1.9 The Present Thesis

This thesis had two goals: 1) to take an evidence-based approach to developing and implementing a BC for novice surgical trainees that is embedded in learning theory and aligns with the most recent shift towards competency-based models of medical education; and 2) explore the longer-term impacts of those changes.

As all components of this thesis revolve around novice skill acquisition and assessment, chapter two provides an overview of the theoretical principles behind novice skill acquisition, combining perspectives from motor learning, cognitive psychology, and education science. This overview includes topics such as Hebbian learning and long-term potentiation, stages of motor learning, working memory, cognitive load theory, divided attention, dual-task conditions, deliberate practice, observational learning, practice schedules and contextual interference, levels of processing, assessment, and feedback. Through the description of how these principles should be integrated within the context of medical education, chapter two provides the theoretical framework for remainder of the thesis.

Following chapter two, this thesis describes three interrelated studies. The first study used a Context, Input, Process, and Product framework to develop and implement a two-week Surgical Foundations BC for novice surgical trainees. Development began in November 2015; teaching sessions were designed to incorporate CBD requirements from the RCPSC, recommendations from staff surgeons, education scientists, and current first- and second-year residents, and various educational principles discussed in chapter two. Development was completed in an iterative manner and the BC was implemented in July 2016, 2017 and 2018. This process is described in chapter three.

The second study compared Objective Structured Clinical Exam (OSCE) performance of surgical trainees who completed traditional training, with surgical trainees that completed the July 2016 BC, at one and two years into training. Results from the OSCE were then integrated with qualitative data, such as focus groups and surveys with trainees, and interviews with Program Directors, to explore BC effectiveness using a Convergent Parallel Mixed Methods (CPMM) approach. This study is described in chapter four.

The third study expanded on the results from chapters three and four to explore how OSCE performance relates to novice trainee performance in the clinical environment. In this study, OSCE performance was compared with performance on workplace-based assessments. This study is described in chapter five.

Chapter six, the final chapter of this thesis, presents a synthesis of the research presented, as well as a discussion surrounding the implications, limitations, and future directions of this work.

1.10 References

Academy of Medical Royal Colleges. (2009). *Common Competencies Framework for Doctors*. Retrieved from https://www.aomrc.org.uk/wp-content/uploads/2018/03/CCFD-August-2009-1.pdf

Accreditation Council for Graduate Medical Education. (1999). *Outcome Project: General Competencies*. Retrieved from https://www.acgme.org/outcome/comp/compFull.asp

Association of American Medical Colleges. (n.d.). *The Road to Becoming a Doctor*. Retrieved from https://www.aamc.org/download/68806/data/road-doctor.pdf

Bell, R. H., Jr., Biester, T. W., Tabuenca, A., Rhodes, R. S., Cofer, J. B., Britt, L. D., & Lewis, F. R., Jr. (2009). Operative experience of residents in US general surgery programs: a gap between expectation and experience. *Ann Surg, 249*(5), 719-724. doi:10.1097/SLA.0b013e3181a38e59

Blackmore, C., Austin, J., Lopushinsky, S. R., & Donnon, T. (2014). Effects of Postgraduate Medical Education "Boot Camps" on Clinical Skills, Knowledge, and Confidence: A Meta-Analysis. *J Grad Med Educ, 6*(4), 643-652. doi:10.4300/JGME-D-13-00373.1

Blackmore, C., Puligandla, P. S., Emil, S., Romao, R., & Lopushinsky, S. R. (2019). A transition to discipline curriculum for pediatric surgery trainees: Evaluation of a pediatric surgery boot camp from 2017 to 2018. *J Pediatr Surg*. doi:10.1016/j.jpedsurg.2019.01.047

Bligh, J. (2002). The first year of doctoring: still a survival exercise. *Medical Education, 36*, 2-3. doi:10.1046/j.1365-2923.2002.01129.x

Block, A. J. (1994). Revisiting the Libby Zion Case. *Chest, 105*(4), 977-978. doi:10.1378/chest.105.4.977

Bohnen, J. D. (2018). A Surgical Transition in Need of Disruption. *Acad Med, 93*, 679-680. doi:10.1097/ACM.0000000000002079

British Medical Association. (2018). *Working in a system that is under pressure*. Retrieved from https://www.bma.org.uk/collective-voice/influence/key-negotiations/nhs-pressures/working-in-a-system-under-pressure

Cameron, A., Millar, J., Szmidt, N., Hanlon, K., & Cleland, J. (2014). Can new doctors be prepared for practice? A review. *The Clinical Teacher, 11*, 188-192. doi:10.1111/tct.12127/suppinfo

Cameron, J. L. (1996). William Stewart Halsted: Our Surgical Heritage. *Ann Surg, 225*, 445-458. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1190776/pdf/annsurg00027-0013.pdf.

Cave, J., Woolf, K., Jones, A., & Dacre, J. (2009). Easing the transition from student to doctor: how can medical schools help prepare their graduates for starting work? *Med Teach, 31*(5), 403-408. doi:10.1080/01421590802348127

Chin, C. J., Roth, K., Rotenberg, B. W., & Fung, K. (2014). Emergencies in otolaryngology-head and neck surgery bootcamp: a novel Canadian experience. *Laryngoscope, 124*(10), 2275-2280. doi:10.1002/lary.24754

Cleland, J., Walker, K. G., Gale, M., & Nicol, L. G. (2016). Simulation-based education: understanding the socio-cultural complexity of a surgical training 'boot camp'. *Med Educ, 50*(8), 829-841. doi:10.1111/medu.13064

Cohen, E. R., Barsuk, J. H., Moazed, F., Caprio, T., Didwania, A., McGaghie, W. C., & Wayne, D. B. (2013). Making July safer: simulation-based mastery learning during intern boot camp. *Acad Med, 88*(2), 233-239. doi:10.1097/ACM.0b013e31827bfc0a

Creswell, J. W. (2007). *Qualitative Research Inquiry Methods; Second Edition*. Thousand Oaks, CA: SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3 ed.). Thousand Oaks, CA: SAGE Publications.

Dobson, J. (1974). Barber into surgeon. *Annals of the Royal College of Surgeons of England, 54*, 84 - 91. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2388370/pdf/annrcse00850-0033.pdf.

Dobson, J., & Walker, R. M. (1979). *Barbers and Barber-Surgeons of London: A History of the Barbers' and Barber-Surgeons Companies*. Osney Mead, Oxford: Blackwell Scientific Publications.

Dorsey, E. R., Nincic, D., & Schwartz, J. S. (2006). An Evaluation of Four Proposals to Reduce the Financial Burden of Medical Education. *Acad Med, 81*, 245-251. Retrieved from https://ovidsp-tx-ovid-com.libaccess.lib.mcmaster.ca/.

Dudek, N., Marks, M. B., & Regehr, G. (2005). Failure to fail: The perspectives of clinical supervisors. *Acad Med, 80*(10), S84-S87. Retrieved from https://ovidsp-tx-ovid-com.libaccess.lib.mcmaster.ca/.

Dyrbye, L., & Shanafelt, T. (2016). A narrative review on burnout experienced by medical students and residents. *Med Educ, 50*(1), 132-149. doi:10.1111/medu.12927

Fernandez, G. L., Page, D. W., Coe, N. P., Lee, P. C., Patterson, L. A., Skylizard, L., . . . Seymour, N. E. (2012). Boot cAMP: educational outcomes after 4 successive years of preparatory simulation-based training at onset of internship. *J Surg Educ, 69*(2), 242-248. doi:10.1016/j.jsurg.2011.08.007

Ferris, H., & Flynn, D. O. (2015). Assessment in Medical Education; What Are We Trying to Achieve? *International Journal of Higher Education, 4*(2), 139-144. doi:10.5430/ijhe.v4n2p139

Flexner, A. (1910). *Medical Education in the United States and Canada: A report to the Carnegie Foundation for the Advancement of Teaching*. New York City, NY: The Carnegie Foundation for the Advancement of Teaching.

Frank, J. R., Snell, L., & Sherbino, J. (2015). *CanMEDS 2015 Physician Competency Framework*. Retrieved from Ottawa:

Frank, J. R., Snell, L., Sherbino, J., & editors. (2015). *CanMEDS 2015 Physician Competency Framework*. Ottawa, ON: Royal College of Physicians and Surgeons of Canada.

Frank, J. R., Snell, L. S., Cate, O. T., Holmboe, E. S., Carraccio, C., Swing, S. R., . . . Harris, K. A. (2010). Competency-based medical education: theory to practice. *Med Teach, 32*(8), 638-645. doi:10.3109/0142159X.2010.501190

Franzese, C. B., & Stringer, S. P. (2007). The evolution of surgical training: perspectives on educational models from the past to the future. *Otolaryngol Clin North Am, 40*(6), 1227-1235, vii. doi:10.1016/j.otc.2007.07.004

Gaskell, N., Hinton, R., Page, T., Elvins, T., & Malin, A. (2016). Putting an end to Black Wednesday: improving patient safety by achieving comprehensive trust induction and mandatory training by day 1. *Clinical Medicine, 16*. doi:10.7861/clinmedicine.16-2-124.

Gawande, A. (2010). *The checklist manifesto: How to get things right*. New York: Metropolitan Books.

Gulland, A. (2017). Is being a doctor bad for your health? *Bmj*. doi:10.1136/bmj.j1803

Hodges, B. D. (2010). A tea-steeping or i-Doc model for medical education? *Acad Med, 85*(9 Suppl), S34-44. doi:10.1097/ACM.0b013e3181f12f32

Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The role of assessment in competency-based medical education. *Med Teach, 32*(8), 676-682. doi:10.3109/0142159X.2010.500704

Imperial College of London. (2009). *New study shows fresh thinking required on week junior doctors start*. Retrieved from https://www.imperial.ac.uk/news/

Jen, M. H., Bottle, A., Majeed, A., Bell, D., & Aylin, P. (2009). Early in-hospital mortality following trainee doctors' first day at work. *PLoS One, 4*(9), e7103. doi:10.1371/journal.pone.0007103

Kilminster, S., Zukas, M., Quinton, N., & Roberts, T. (2011). Preparedness is not enough: understanding transitions as critically intensive learning periods. *Med Educ, 45*(10), 1006-1015. doi:10.1111/j.1365-2923.2011.04048.x

Lerner, B. H. (2009). Life-changing case for doctors in training. *New York Times*.

McMaster University, D. o. S. (2017). Surgical Foundations. *Education.* Retrieved from https://surgery.mcmaster.ca/education/surgical-foundations

McQueen, S. A., Petrisor, B., Bhandari, M., Fahim, C., McKinnon, V., & Sonnadara, R. R. (2016). Examining the barriers to meaningful assessment and feedback in medical training. *Am J Surg, 211*(2), 464-475. doi:10.1016/j.amjsurg.2015.10.002

National Health Service Employers. (2009). *Working Time Directive: FAQs for trust implementation teams*. Retrieved from https://www.nhsemployers.org/about-us/nhs-european-office/working-time-directive

National Steering Committee on Resident Duty Hours. (2013). *Fatigue, Risk, & Excellence: Towards a Pan-Canadian Consensus on Resident Duty Hours*. Retrieved from http://www.residentdutyhours.ca

New Zealand Resident Doctors' Association. (2007). *New Zealand Resident Doctors' Association Collective Agreement*. Retrieved from https://www.nzrda.org.nz

Parent, R. J., Plerhoples, T. A., Long, E. E., Zimmer, D. M., Teshome, M., Mohr, C. J., . . . Dutta, S. (2010). Early, intermediate, and late effects of a surgical skills "boot camp" on an objective structured assessment of technical skills: a randomized controlled study. *J Am Coll Surg, 210*(6), 984-989. doi:10.1016/j.jamcollsurg.2010.03.006

Patel, N. (2014). Learning lessons: the Libby Zion case revisited. *J Am Coll Cardiol, 64*(25), 2802-2804. doi:10.1016/j.jacc.2014.11.007

Patel, R., Drover, A., & Chafe, R. (2015). Pediatric faculty and residents perspectives on In Training Evaluation Reports (ITERs). *Canadian Medical Education Journal, 6*(2), e41-e53. Retrieved from http://www.cmej.ca.

Pattani, R., Wu, P. E., & Dhalla, I. A. (2014). Resident duty hours in Canada: past, present and future. *CMAJ, 186*(10), 761-765. doi:10.1503/cmaj.131053

Peyre, S. E., Peyre, C. G., Sullivan, M. E., & Towfigh, S. (2006). A surgical skills elective can improve student confidence prior to internship. *J Surg Res, 133*(1), 11-15. doi:10.1016/j.jss.2006.02.022

Phillips, D. P., & Barker, G. E. (2010). A July spike in fatal medication errors: a possible effect of new medical residents. *J Gen Intern Med, 25*(8), 774-779. doi:10.1007/s11606-010-1356-3

Prince, K., Van de Wiel, M., Van der Vleuten, C., Boshuizen, H., & Scherpbier, A. (2004). Junior doctors' opinions about the transition from medical school to clinical practice: a change of environment. *Educ Health (Abingdon), 17*(3), 323-331. doi:10.1080/13576280400002510

Qureshi, Z., & Maxwell, S. (2012). Has bedside teaching had its day? *Adv Health Sci Educ Theory Pract, 17*(2), 301-304. doi:10.1007/s10459-011-9308-1

Riebschleger, M., & Philbert, I. (2011). *The ACGME 2011 Duty Hour Standards*. Retrieved from https://www.acgme.org/

Robinson, J. O. (1984). The Barber-Surgeons of London. *Arch Surg, 119*, 1171 - 1175. Retrieved from https://jamanetwork.com/.

Roland, C. G. (1982). *Sir William Osler 1849 - 1919. A Selection for Medical Students*: Associated Medical Services Inc.

Royal College of Physicians and Surgeons of Canada. (2010). *Objectives of Surgical Foundations Training*. Retrieved from http://www.royalcollege.ca/

Royal College of Physicians and Surgeons of Canada. (2016). Competence by Design Cheat Sheet. Retrieved from www.royalcollege.ca

Royal College of Physicians and Surgeons of Canada. (2018). CBD Start, Launch and Exam Schedule. Retrieved from http://www.royalcollege.ca/.

Royal College of Physicians and Surgeons of Canada. (2019). CBD Implementation. Retrieved from http://www.royalcollege.ca/rcsite/cbd/cbd-implementation

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144. doi:10.1007/BF00117714

Safir, O., Sonnadara, R., Mironova, P., & Rambani, R. (2018). *Boot Camp Approach to Surgical Training*: Springer International Publishing.

Satterthwaite, L., Leighton, J., Safir, O., & editors. (2018). Surgical Preparatory Camp (PREP Camp). In *Boot Camp Approach to Surgical Training*: Springer International Publishing.

Shanafelt, T. D., Hasan, O., Dyrbye, L. N., Sinsky, C., Satele, D., Sloan, J., & West, C. P. (2015). Changes in Burnout and Satisfaction With Work-Life Balance in Physicians and the General US Working Population Between 2011 and 2014. *Mayo Clin Proc, 90*(12), 1600-1613. doi:10.1016/j.mayocp.2015.08.023

Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., . . . Dugravot, A. (2012). Timing of onset of cognitive decline: results from Whitehall II prospective cohort study. *Bmj, 344*, d7622. doi:10.1136/bmj.d7622

Sonnadara, R., McQueen, S., Mironova, P., Nousiainen, M., Ferguson, P., Alman, B., . . . Reznick, R. (2013). Reflections on current methods evaluating skills during joint replacement surgery. *The Bone & Joint Journal, 95-B*(11), 1445-1449. doi:10.1302/0301-620X.95B11

Sonnadara, R. R., Garbedian, S., Safir, O., Nousiainen, M., Alman, B., Ferguson, P., . . . Reznick, R. (2012). Orthopaedic Boot Camp II: examining the retention rates of an intensive surgical skills course. *Surgery, 151*(6), 803-807. doi:10.1016/j.surg.2012.03.017

Sonnadara, R. R., Van Vliet, A., Safir, O., Alman, B., Ferguson, P., Kraemer, W., & Reznick, R. (2011). Orthopedic boot camp: examining the effectiveness of an intensive surgical skills course. *Surgery, 149*(6), 745-749. doi:10.1016/j.surg.2010.11.011

Strohnbehn, G. W., Cronin, D. T., & Levy, K. (2018). Are They Ready to Be Senior Residents? *Acad Med, 93*, 679. doi:10.1097/ACM.0000000000002079

Sullivan, R. (1987). Grand Jury Assails Hospital in /84 Death of 18-year-old. *The New York Times*.

Swadron. (2018). Serotonin Syndrome and the Libby Zion Affair. *Emergency Physicians Monthly*. Retrieved from http://epmonthly.com/article/serotonin-syndrome-and-the-libby-zion-affair/

Teagle, A. R., George, M., Gainsborough, N., Haq, I., & Okorie, M. (2017). Preparing medical students for clinical practice: easing the transition. *Perspect Med Educ, 6*(4), 277-280. doi:10.1007/s40037-017-0352-2

ten Cate, O. (2017). Competency Based Postgraduate Medical Education: Past Present and Future. *GMS Journal for Medical Education, 34*, 1-13. doi:10.3205/zma001146

ten Cate, O. (2018). A primer on entrustable professional activities. *Korean J Med Educ, 30*(1), 1-10. doi:10.3946/kjme.2018.76

Teo, A. R., Harleman, E., O'Sullivan P, S., & Maa, J. (2011). The key role of a transition course in preparing medical students for internship. *Acad Med, 86*(7), 860-865. doi:10.1097/ACM.0b013e31821d6ae2

Teunissen, P. W., & Westerman, M. (2011). Opportunity or threat: the ambiguity of the consequences of transitions in medical education. *Med Educ, 45*(1), 51-59. doi:10.1111/j.1365-2923.2010.03755.x

van der Vleuten, C. P., Schuwirth, L. W., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Med Teach, 34*(3), 205-214. doi:10.3109/0142159X.2012.652239

Young, J. Q., Ranji, S. R., Wachter, R. M., Lee, C. M., Niehaus, B., & Auerbach, A. D. (2011). "July Effect": Impact of Academic Year-End Changeover on Patient Outcomes. *Annals of Internal Medicine, 155*, 309-315. Retrieved from https://www.annals.org/.

Chapter 2

Bridging the Gap: Theoretical Principles Behind Surgical Boot Camps

Citation:

Wagner, N., McQueen, S., and Sonnadara, R.R. (2018). "Bridging the gap: theoretical principles behind surgical boot camps". R. Rambani (Ed.) *Boot Camp Approaches to Surgical Training.* (pp. 1-11). Basel, Switzerland: Springer.

2.0 Bridging the Gap: Theoretical Principles Behind Surgical Boot Camps

2.1 Introduction

In the past few years, motor behaviorists, cognitive psychologists, and educational scientists have made significant contributions to understanding the processes involved in skill acquisition and learning. Though these fields use different terminology and typically discuss different theories, frequently they are talking about similar concepts from different perspectives. Exploring the intersection between these fields will provide a broader view on how novices acquire new skills, and can help medical educators structure boot camps (BCs) to promote effective and efficient learning at the onset of residency.

2.2 Theoretical Principles of Novice Skill Acquisition

2.2.1 Motor Programs

In the late 1800s, Robert S. Woodworth proposed that psychologists should study relationships between the stimulus, response, and organism (Woodworth, 1899). Woodworth's idea that these three components were interconnected, and that the environment does things to the individual and the individual does things to the environment, was heavily influenced by his work with his doctoral supervisor, Raymond Cattell, and was groundbreaking at the time. Namely, Woodworth proposed the concept of 'sets' (Benjafield, 2015). To Woodworth, a set was a temporary organization of the brain that facilitated some responses and inhibited others (Benjafield, 2015). Following Woodworth's hypotheses, Karl Lashley, another of Cattell's doctoral students, conducted a series of experiments, which confirmed that goal-directed movements result from a chain of sensory-motor reactions due to some form of neural organization (Lashley, 1929). In 1963, Steven W. Keele described the chain of sensory-motor reactions as a 'motor program'. Keele suggested that a motor program, or pre-defined movement pattern, existed for every individual action (Summers & Anson, 2009). However, Keele's description opened the door for debate.

In 1967, Nikolai Bernstein identified the degrees of freedom issue in motor programs, stating there are far too many individual parts of a movement for the brain to control each aspect separately on a conscious level. Following Bernstein's comment, two major problems with motor programs were recognized: a) storage - it seemed impossible that the brain could store all possible combinations of every movement; and b) novelty- this theory did not address how an individual creates a new motor program (Schmidt & Lee, 2005). Identification of the shortcomings with this theory led to the creation of Jack Adams' closed-loop model of motor programs, which postulated motor programs existed for groups of similar movements rather than every individual movement (Adams, 1971). Adams' theory proposed that learners were able to select the appropriate motor program, detect errors and refine movement through feedback. This was promising - however, Adams' closed-loop theory focused on slow, purposeful movements and lacked evidence supporting fast, more reflex-like movements (Schmidt, 1975).

In response to the gaps in Adams' model, Schmidt (1975) proposed schema theory to explain all movements. Schmidt suggested a 'motor schema' represents a set of rules that determine which movement will be selected based on the initial conditions of the environment and body, the response specifications of the movement, the sensory feedback from the movement, and the movement outcomes (Schmidt, 1975). The initial conditions and response specifications select the appropriate motor schema, while sensory feedback and movement outcomes refine the motor schema for subsequent performances (Schmidt, 1975). With this, Schmidt (1975) proposed that motor schemas could be strengthened through practice, causing muscles to contract and release more efficiently; thus resulting in more consistent, predictable and smooth movements.

2.2.2 Hebbian Learning and Long-Term Potentiation

At the neural substrate level, Hebbian learning principles and long-term potentiation can help explain how practice strengthens motor schemas. Donald Hebb (1949) introduced the concept of "cells that fire together, wire together". According to Hebb's theory, when an axon of cell A repeatedly excites cell B, metabolic changes can occur in one or both of the cells, which increases the synaptic strength between neurons, or communication pathway between neurons. This phenomenon is referred to as long-term potentiation (Shors & Matzel, 1997). In terms of motor behaviour, when a motor schema is selected for any given movement, a group of cells (group A) is activated. Group A then activates the group of cells responsible for movement (group B). The repeated firing of group A and B together can lead to long-term potentiation, or improved communication between neurons and more efficient movement (Hebb, 1949; Shors & Matzel, 1997).

2.2.3 Stages of Motor Learning

Motor learning theories also describe the refinement of movement through practice. In the first stage of motor learning, all movements are consciously controlled as the learner is trying to determine what needs to be done to accomplish the action in a step-by-step fashion (Adams, 1971; Fitts & Posner, 1967). Novice learners often use self-talk, among other strategies, to guide themselves through each step of the desired movement (Schmidt & Wrisberg, 2004). Thus, this stage requires a large attentional capacity, and movements occur in a fairly slow, inaccurate, and inefficient manner (Wulf, 2007). This stage is known as the ‘cognitive stage’ (Fitts & Posner, 1967) or the ‘verbal stage’ (Adams, 1971). According to the model by Fitts and Posner (1967), the next stage called the ‘associative stage’ is where learners use feedback to make subtle movement adjustments and improve performance on subsequent trials. In this stage, some aspects of the movement become more efficient and require less attention (Wulf, 2007). The last stage, referred to as either the ‘autonomous stage’ (Fitts & Posner, 1967) or the ‘motor stage’ (Adams, 1971), is reached when the learner is able to perform movements in a seemingly effortless manner and dedicate attentional resources elsewhere. Though the two models discussed use different terminology, both recognize the increase in performance and decrease in attentional load as learners progress from novice to expert. The importance of attention to motor learning (particularly for novices) is discussed in the Psychology literature on working memory.

2.2.4 Working Memory

As previously mentioned, someone in the first stage of motor learning focuses attention on a task in a step-by-step manner. Focusing on what the body has to do in order to perform the skill requires a significant amount of short-term memory capacity (Fitts & Posner, 1967). Literature refers to short-term memory as working memory, since it contains the ideas or information one is thinking about, or “working on”, right now (E. K.

Miller & Buschmann, 2015; Reisberg, 2013). Unfortunately, according to G. A. Miller (1956) humans have a limited working memory capacity of approximately seven (plus or minus two) blocks, or chunks, of information. According to G. A. Miller (1956), a chunk does not hold a fixed quantity of information, rather is highly varied. Thus, a chunk could be one digit (e.g. the number six), or an entire phone number; all chunks are made on an individual basis (G. A. Miller, 1956). When learning a new motor task, each individual step in the movement would be considered a 'chunk'; thus, depending on the complexity, working memory is often at its capacity when learning novel tasks. As the learner practices the movement and progresses from novice to expert, the individual steps of the action are combined into larger chunks, and thus there is more room in working memory for other information (Norman, Brooks, & Allen, 1989). Once a learner reaches a high level of performance, or the autonomous stage of motor learning, it is possible the entire movement has been condensed into one chunk of information, and therefore little working memory is required to perform the action. Thus, until learners reach the autonomous stage, it is unlikely they have the capacity to handle any additional information to the task being learned.

2.2.5 Cognitive Load Theory

Due to the limited capacity of working memory, careful attention must be paid to the information provided to learners when they are attempting to acquire or perform a new skill. According to the cognitive load theory, the information brought into working memory can be separated into three categories (Sweller, van Merrienboer, & Paas, 1998). Intrinsic load is the information required to learn the task (varies with task complexity);

germane load is the information required to process that new information into a schema (also varies with task complexity); and extraneous load is any other information that negatively interferes with the integration of the intrinsic information (Sweller et al., 1998). Since novice learners dedicate the majority of their attentional resources to intrinsic information, it is important to minimize the extraneous load, or information not relevant to the task at hand. Even if the information is equally important, such as crucial information about a patient within the context of healthcare, novice learners do not have the capacity to attend to a novel task *and* the extraneous information at the same time. It is therefore important that educators are sensitive to *when* information is presented to learners.

2.2.6 Divided Attention

Multi-tasking, or dividing attention, can have detrimental effects on memory and performance in learners (Baddeley, 1986; F. I. M. Craik & Lockhart, 1972). Since working memory has a limited capacity, attending to more than one task can result in a competition for attentional resources. Accordingly, an individual can only perform multiple tasks simultaneously if they have enough cognitive capacity, or resources, to do so. If the combined resources required for multiple tasks exceed the working memory capacity, the individual must switch back and forth between tasks (F. I. Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; E. K. Miller & Buschmann, 2015; Reisberg, 2013). Switching back and forth between tasks may cause the learner to miss vital information, negatively affecting performance. In other words, if an instructor attempts to give information to a novice learner while they are performing a task, performance on the original task will decrease and it is unlikely the learner will be able to recall the additional information provided by the instructor.

2.2.7 Dual-task Conditions

Many studies have investigated the difference in the novice versus expert ability to divide attention in dual-task conditions. Beilock, Wierenga, and Carr (2002) conducted a golf-putting study, which recorded novices' and experts' ability to putt while performing an auditory word search task. Experts had a higher recognition memory for the words presented to them while putting, and a diminished memory of the putt itself when compared to novices. Since experts typically automatize actions into fewer chunks of information (Norman et al., 1989), they require little attentional resources to complete a task such as golf putting. As a result, experts have the working memory capacity to attend to additional information, whereas novices do not.

Dubrowski, Brydges, Satterthwaite, Xeroulis, and Classen (2012) performed a similar comparison using dual-task conditions that often occur in medical training. Their paper entitled 'Do not teach me while I am working!" discusses how medical trainees are responsible for learning both technical and nontechnical skills, and often the teaching of the two occurs simultaneously in the clinic and operating room environments. As novice trainees dedicate the majority of their attentional resources to the first task (e.g. the technical skill), they do not have the working memory capacity to take in any additional information (e.g. information on the nontechnical skill). Thus, attempting to teach both skills simultaneously does not promote more efficient learning, but rather decreases performance and memory on all tasks. In order to ensure novice trainees have the opportunity to attend to all of the information, educators must take steps to avoid placing novices in dual-task conditions.

2.2.8 Interim Summary

A variety of constraints exist when attempting to learn a novel task. In the earliest stages of motor learning, novices require a significant amount of attentional resources to execute the task in a step-by-step manner. Due to the limited capacity of working memory, novices are unable to divide their attention and complete additional tasks without negatively affecting performance. However, as learners progress through the stages of motor learning, the attentional load required to perform the main task decreases and their attention may be dedicated to other tasks. Therefore, educators need to be mindful that different stages of learning require different teaching methods for effective long-term learning.

2.3 Practice Makes ~~Perfect~~ Permanent

The previous section reviews the theoretical principles guiding how novice learners can improve performance and decrease the attentional load with practice, bringing the phrase "practice makes perfect" to mind. However, motor learning literature suggests the organization of a practice session heavily influences whether a learner will have a temporary increase in performance or a long-term increase in performance (i.e. learning). Thus, a more accurate phrase for motor learning may be "perfect practice makes permanent" (Gladwell, 2008). Based on motor behaviour, cognitive science, and education literature, this section reviews how to structure a practice session to promote permanent learning.

2.3.1 Deliberate Practice

Ericsson, Krampe, and Tesch-Romer (1993) suggest that experience alone is not enough to achieve perfection; rather, it is the result of many hours of deliberate practice. Ericsson et al. (1993) define deliberate practice as activities that are specifically focused on improving some aspect of performance. In order for deliberate practice to be most effective, it must include specific and immediate feedback on how to improve performance (Ericsson, 2004). Furthermore, Ericsson and Lehman (1996) suggest that depending on the quality of each practice session, the highest level of human performance requires approximately 10,000 hours of deliberate practice in which learners are constantly placed in training situations where their goals exceed their current level of performance (Ericsson, 2004). Although achieving perfection, or expertise, may not be the goal of residency training, incorporating deliberate practice into trainees' training schedules can greatly increase the quality of performance and the development of expertise in the future (Keith & Ericsson, 2007).

2.3.2 Observational Learning

Often novices like to watch others perform a novel task – does that count as deliberate practice? Many have investigated this process and have suggested that the observation of others, especially when combined with physical practice, can make important and unique contributions to learning (Shebilske, Regian, Arthur, & Jordan, 1992). Neuroimaging studies suggest an overlap exists between the neurons activated in both action production and observation of movement (Gallese, Fadiga, Fogassi, & Rizzolati, 1996). These neurons are referred to as mirror neurons (Gallese et al., 1996;

Rizzolati & Craighero, 2004). Lago-Rodriguez, Lopez- Alonso, and Fernandez-del-Olmo (2013) suggest that activation of mirror neurons during observation allows learners to build more accurate estimates of performance outcomes for a practiced task. Moreover, observational learning may remove the attentional load associated with movement production and allow novice learners to attend to additional information on performance outcomes (C. H. Shea, Wulf, & Whitacre, 1999). This additional information can then be encoded to strengthen the pre-existing schema. Therefore, incorporating the combination of action and observation in practice sessions can serve as a strategy to overcome the effects of the limited capacity of working memory on novice skill acquisition.

2.3.3 Practice Schedules

Motor behaviourists have long studied what types of (deliberate) practice are most effective for learning. The literature suggests intermixing practice with periods of rest or other tasks (distributed practice) promotes skill acquisition and retention in comparison with a continuous block of practice (massed practice) (Lee & Genovese, 1988; Schmidt & Bjork, 1992). For example, Moulton et al. (2006) investigated the influence of massed versus distributed practice on surgical skills. Participants were split into two groups; the massed group received four consecutive training sessions in one day, while the distributed group received the same training sessions over four weeks (one session per week). Both groups showed immediate improvements in performance, but the distributed group performed significantly better on all retention tests. Therefore, this study by Moulton et al. (2006), among others, suggests distributed practice of skills promotes better retention and transfer of learning.

2.3.4 Contextual Interference

One important difference between massed and distributed practice is the amount of contextual interference (Lee, Wishart, Cunningham, & Carnahan, 1997). Contextual interference was originally studied by Battig (1966) and refers to any aspect of the practice session that interferes with learning. In a massed practice session, the same task is repeatedly practiced and contextual interference is low, while in a distributed practice session the learner switches between tasks and the contextual interference is high (J. B. Shea & Morgan, 1979). There are two principal explanations for contextual interference, the elaborative processing hypothesis and the forgetting-reconstruction hypothesis. The elaborative processing hypothesis suggests that distributed practice forces the learner to make inter-task comparisons and connections (J. B. Shea & Titzer, 1993). According to the levels of processing theory, forming connections between new information and previous knowledge results in deeper processing (F. I. M. Craik & Lockhart, 1972). Deeper processing then leads to more connections to the same memory representation, making it easier to retrieve information from long-term memory (F. I. M. Craik & Lockhart, 1972). In comparison, the forgetting-reconstruction hypothesis suggests that distributed practice forces the learner to “forget” the motor program for the original task and abandon it from working memory to make room for information on the new task. As a result, the learner has to reconstruct (or retrieve) the motor program from long-term memory every time they return to the original task (Lee & Magill, 1983). Every time the schema is reconstructed from long-term memory, the retrieval path is strengthened.

Therefore, though the two theories highlight different aspects, they both emphasize the importance of retrieving information from long-term memory.

2.3.5 Encoding and Retrieval

There are many analogies to help explain the importance of encoding and retrieval for learning. Information comes into working memory (the inbox on a desk), and then gets processed and stored, or encoded, into long-term memory (the office filing cabinet) (Higbee, 1996). As previously mentioned, the levels of processing theory suggests assigning meaning to new information, or building connections between new and old information, promotes deeper processing (F. I. M. Craik & Lockhart, 1972). If deeper processing occurs, it is easier to retrieve information from long-term memory in the future. In other words, if a new file is placed with old files on a similar topic, it will be easier to find. The more connections to that information (e.g. knowing the group of files that go in front or behind) increases the likelihood of remembering the content and where it is stored, resulting in quicker and more accurate retrieval. Thus, for novice learners, it is important to link new motor tasks or information with existing schemas. This will allow the novel information to be integrated with prior knowledge, requiring less cognitive effort for encoding and retrieval (van Kesteren, Rijpkema, Ruiter, Morris, & Fernandez, 1975).

2.3.6 Assessment for Learning

Another way to promote retrieval is through repeated assessment. The act of retrieving the correct information from long-term memory repeatedly activates the neurons associated with the correct response and makes it easier to retrieve in the future

(Roediger, Putnam, & Smith, 2011). However, often learners are unable to recall the information successfully. Does this still have a learning benefit? Consider the phrase "…won't make that mistake again"; making an incorrect response, especially on a test, will usually create a strong memory (Giordana & Serra, 2001). This strong memory can act as a trigger, building a connection (or retrieval path) back to the correct response (F. I. M. Craik & Lockhart, 1972). Thus, regardless of whether learners answer correctly, testing (and retesting) can often be more powerful than re-studying information (Roediger et al., 2011). Moreover, some assessments are more beneficial for learning than others.

As mentioned in chapter one, there are two types of assessment, summative and formative. The primary difference between the two is the purpose and the consequence of the assessment (Sadler, 1989). Summative assessment attempts to summarize the achievements of a learner, usually for a final grade or certification (Sadler, 1989). Formative assessment is often used to evaluate how performance outcomes can be used to improve overall learning (Sadler, 1989). Using formative assessment to provide learners with feedback regarding gaps between their actual performance and the desired performance is a crucial piece of learning (Ramprasad, 1983; Sadler, 1989).

2.3.7 Feedback

Feedback was one of the key components of both Adams' (1971) and Schmidt's (1975) theories of motor programs and schemas discussed in the beginning of this chapter (2.2.1). In motor learning, feedback is typically broken down into two categories. Knowledge of performance refers to the kinematic, or sensory feedback the learner receives when they perform an action (Schmidt & Wrisberg, 2004), whereas knowledge of results refers to extrinsic information provided to the learner regarding their success in accomplishing the goal (Salmoni, Schmidt, & Walter, 1984). From a motor learning perspective, these two types of feedback can be used to refine and strengthen a motor program or schema. From a cognition standpoint, feedback is considered a valuable tool to help promote deep processing (F. I. M. Craik & Lockhart, 1972). From an educational perspective, formative assessment can be used to help learners understand the overall goal, compare their performance against that goal, and gain valuable feedback on how to bridge the gap between the two (Sadler, 1989). Thus, although each field looks at feedback through a slightly different lens, they all acknowledge the critical role of feedback in improving performance and learning.

As a result, researchers from many domains have studied what makes the most effective feedback (Janelle, Kim, & Singer, 1995). Some suggest that specific and immediate feedback should be given after every instance of deliberate practice (Ericsson, 2004), while others suggest that self-regulated feedback, or feedback controlled by the learner, can result in significant improvements in performance compared with groups where the instructor controls feedback (Chiviacowsky & Wulf, 2002). Self-regulated feedback forces learners to be actively involved in the learning process, thus promoting deeper processing (Watkins, 1984), more accurate error estimation (Chiviacowsky & Wulf, 2002), and increased motivation (Bandura, 1993). The idea that increased autonomy may be beneficial for performance is true for not only feedback, but setting practice schedules as well (Safir, Williams, Dubrowski, Backstein, & Carnahan, 2013). Allowing learners to decide what they want to practice (within specific guidelines), and when they want feedback gives them the opportunity to tailor the practice session to their individual needs, resulting in more meaningful practice (Safir et al., 2013).

Other researchers have explored what type of language is most effective when giving learners feedback. As discussed in 2.2.1, learners tend to focus on what their body needs to do in a step-by-step manner in the first stage of motor learning. This is considered an intrinsic focus of attention. As learners progress, less attention is required to perform the task, and they shift their attention to extrinsic factors (Wulf, Hob, & Prinz, 1998). For example, a novice golfer focuses attention on what their arms are doing during a swing; yet an expert golfer may focus on the speed of the club, or direction of the ball. Wulf et al. (1998) were amongst the first to examine the effect of intrinsic versus extrinsic attention on performance using a ski simulator. One group of participants was instructed to focus on their outer foot as the ski simulator moved; when the simulator moved left, they had to push with their left foot, and when it moved to the right, they had to push with their right foot; this represented an intrinsic focus. The other group was instructed to focus on the wheels of the platform (which were located right under their feet); this represented an extrinsic focus. The results demonstrated that an extrinsic focus improved both immediate performance and retention compared with the other two groups. Thus, this study by Wulf et al. (1998) was the first to show that shifting attention to extrinsic factors could actually improve performance. Indeed, an extrinsic focus may decrease the attentional demands of a task and therefore promote more automatic control and improved performance (Wulf, McNevin, Fuchs, Ritter, & Toole, 2000; Wulf, McNevin, Shea, & Wright, 2001). This highlights that the type of language used to instruct novice learners is important and can be manipulated to promote automaticity and improved performance (Wulf et al., 2000).

2.3.8 Interim Summary

There are many factors to be considered when trying to determine how a practice session may best be structured to promote learning in novices. Deliberate, or reflective, practice can greatly improve performance and learning. Observational learning can provide learners with information that they may have missed when performing the task themselves. Practice sessions arranged in a distributed manner promote the establishment of connections and deep processing. Frequent formative assessments can be used to strengthen those connections and retrieval paths. Allowing learners to control the timing and frequency of feedback used in formative assessment and their practice schedules can further improve learning. Lastly, instructions and feedback that emphasize an extrinsic focus of attention can result in faster automation and improved performance. Thus, considering the theoretical foundations underpinning how novices learn new skills can play a vital role in the effectiveness of training programs in any domain.

2.4 Creating an Effective Boot Camp

Understanding the theoretical foundations for what training methods are most effective for novice learners can greatly assist in the development of a successful boot camp (BC) for trainees during the transition between medical school and residency. A BC program should give learners the opportunity to participate in many hours of deliberate practice in an environment where they are able to safely make mistakes and learn from them, without compromising patient care. Although it is impossible for new trainees to attain expertise within a BC (recall Ericsson's 10,000 hour suggestion), partaking in deliberate practice, or practice with the goal of improving a specific aspect of performance, can improve immediate performance and set the standard for work ethic and life-long learning from the onset of residency (Keith & Ericsson, 2007). In order to facilitate deliberate practice, the goals and objectives of each BC session should be outlined and presented to trainees to guide their learning. After the goals and objectives have been identified, the remainder of the educational activity (i.e. assessments and teaching strategies) should be considered. This is often referred to as constructive or "backwards" alignment (Biggs, 1996).

When considering the most effective teaching strategies, trainees should be given the opportunity to participate in deliberate practice through performing the task, or observing someone else perform the task, when possible. The use of both physical practice and observational learning will allow trainees to build and refine motor schemas for the skills, and also encode any relevant information that they may have missed during their own physical practice (recall the limited capacity of working memory). Incorporating both observation and physical practice may be accomplished through instructor demonstrations and practice periods, or working in small peer groups.

Secondly, effective BCs should include distributed practice sessions. Whether trainees are observing or performing the skill, interspersing the practice of various skills and rest periods will force trainees to repeatedly retrieve the new information from long-term memory. The more often trainees retrieve information, the more likely they are to remember the skills and transfer their knowledge to clinical practice. Consider the filing cabinet analogy - the more often the file is retrieved and reviewed, the more likely the location and content will be remembered.

Incorporating feedback and assessment is another critical component of successful BCs. Both deliberate practice (Ericsson, 2004) and the strengthening of motor schemas (Adams, 1971; Schmidt, 1975) require immediate and specific feedback for deep processing. This feedback can be provided to trainees through formative assessment, which aims to identify the goal of the task, compare trainees' performance to the goal, and offer guidance on how to bridge any gaps between the two (Sadler, 1989). Repeated formative assessments during a BC also force trainees to practice retrieving new information, and whether they answer correctly or not, feedback from the assessment will promote deeper processing and retention. Formative assessment can be built into each session through instructor-based feedback and/or peer-based feedback.

Instructors must also be aware of their timing and language choices when giving trainees feedback. In the beginning stages of motor learning, attention is focused on completing the task; thus, if an instructor provides a novice trainee with feedback while they are performing the task, trainees may not be able to encode that information. Therefore, feedback should be provided after a task, or when trainees ask for it (Chiviacowsky & Wulf, 2002). Furthermore, when providing trainees with feedback, instructors should use language focused around extrinsic factors rather than intrinsic factors (e.g. giving feedback on the position of the needle during a suturing task rather than position of fingers). This has the potential to help trainees progress to the automatic stage of motor learning faster by decreasing the amount of cognitive resources required and increasing their ability to add secondary tasks.

Additionally, gauging trainees' skill level prior to instruction can ensure appropriate teaching. For this purpose, incorporation of the briefing, intraoperative teaching, and debriefing (BID) method could be valuable in BCs. Briefing refers to a short interaction prior to the activity where the learner can assess their needs and jointly establish learning objectives with the instructor (Roberts, Williams, Kim, & Dunnington, 2009). This step will assist the instructor in identifying what stage the learner is at, and what information is appropriate to promote learning. The second step, intraoperative teaching, refers to focusing instruction on the specific shared learning goals (Roberts et al., 2009). This will lessen any extraneous information that may be detrimental to learning and overload the learners' attentional resources. The last step, debriefing, is comprised of four components: reflection by the learner, rules for future attempts, reinforcement of what was done right, and correction of any errors (Roberts et al., 2009). This step emphasizes feedback and formative assessment as a critical component for improving future attempts.

2.5 Summary

Boot camps at onset of residency will contain a multitude of novel tasks that trainees must learn. In order to effectively teach those novel tasks, medical educators must consider what training methods are best for novice skill acquisition. To facilitate permanent learning, BCs must involve deliberate practice, constructive alignment, observational learning, distributed practice, frequent formative assessments, self-regulated feedback, and instruction focused towards extrinsic factors. Developing a BC founded in these principles offers an opportunity for trainees to improve their medical knowledge, confidence, and procedural and technical skills in a safe environment. Gaining this valuable exposure prior to starting to care for patients may decrease trainee feelings of unpreparedness as well as patient complication rates in the month of July. Furthermore, acquiring some basic skills at the onset of residency is likely to accelerate progression through subsequent stages of training since less time and attention need to be dedicated to those basic skills. Thus, if BCs are aligned with the theoretical foundations of novice skill acquisition, this type of training paradigm has the potential to revolutionize medical training.

2.6 References

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of Motor Behavior, 3*, 111-150. doi:10.1080/00222895.1971.10734898

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Claredon.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117-148.

Battig, W. F. (1966). Facilitation and interference. In E. A. Bilodeau (Ed.), *Acquisition of Skill*. New York, New York: Academic Press.

Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *The Quarterly Journal of Experimental Psychology Section A, 55*, 1211-1240.

Benjafield, J. G. (2015). *A History of Psychology (4th Edition)*: Oxford University Press.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-367.

Chiviacowsky, S., & Wulf, G. (2002). Self-controlled feedback: Does it enhance learnng because performers get feedback when they need it? *Research Quarterly for Exercise and Sport, 73*, 408-415.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology, 125*, 159-180.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*, 671-684.

Dubrowski, A., Brydges, R., Satterthwaite, L., Xeroulis, G., & Classen, R. (2012). Do not teach me while I am working! *Am J Surg, 203*(2), 253-257. doi:10.1016/j.amjsurg.2010.08.020

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med, 79*, S70.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review, 100*(3), 363-406.

Ericsson, K. A., & Lehman, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaption to task constraints. *Annual Review of Psychology, 47*, 273-305.

Fitts, P. M., & Posner, M. I. (1967). *Human Performance*. Belmont, CA: Brooks/Cole.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolati, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

Giordana, A., & Serra, A. (2001). Learning from mistakes. *Human and Machine Perception, 3*, 89-102.

Gladwell, M. (2008). *Outliers: The story of success* (1st ed.). New York, New York: Little, Brown and Company.

Hebb, D. O. (1949). *The Organization of Behaviour: A neuropsychological theory*. New York, New York: John Wiley & Sons Inc.

Higbee, K. L. (1996). *Your memory: how it works and how to improve it* (2nd ed.). Oxford, England: Da Capo Press.

Janelle, C. M., Kim, J., & Singer, R. N. (1995). Subject-controlled performance feedback and learning of a closed motor skill. *Perceptual and Motor Skills, 81*, 627-634.

Keith, N., & Ericsson, K. A. (2007). A deliberate practice account of typing proficiency in everyday typists. *Journal of Experimental Psychology, 13*, 135-145.

Lago-Rodriguez, A., Lopez- Alonso, V., & Fernandez-del-Olmo, M. (2013). Mirror neuron system and observational learning: Behavioural and neurophysiological evidence. *Behavioural Brain Research, 248*, 104-113. doi:10.1016/j.bbr.2013.03.033

Lashley, K. S. (1929). *The foundations of experimental psychology*. Worchester, MA: Clark University Press.

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, 60*, 59-65.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 730-746.

Lee, T. D., Wishart, L. R., Cunningham, S., & Carnahan, H. (1997). Modeled timing information during random practice eliminates and contextual interference effect. *Research Quarterly for Exercise and Sport, 68*, 100-105.

Miller, E. K., & Buschmann, T. J. (2015). Working memory capacity: limits on the bandwidth of cognition. *Daedalus, 144*, 112-122.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*, 81-97.

Moulton, C.-A. E., Dubrowski, A., Macrae, H., Graham, B., Grober, A., & Reznick, R. (2006). Teaching surgical skills: What kind of practice makes perfect? *Annals of Surgery, 244*, 400-409.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1166-1174.

Ramprasad, A. (1983). On the definition of feedback. *Behavioural Science, 28*, 4-13.

Reisberg, D. (2013). *Cognition Exploring the Science of the Mind* (6th ed.). New York, New York: Norton & Company.

Rizzolati, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Roberts, N. K., Williams, R. G., Kim, M. J., & Dunnington, G. L. (2009). The briefing, intraoperative teaching, debriefing model for teaching in the operating room. *Am Coll Surg, 208*, 299-303.

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). *Ten benefits of testing and their applications to educational practice*. Oxford, England: Elsevier.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144. doi:10.1007/BF00117714

Safir, O., Williams, C. K., Dubrowski, A., Backstein, D., & Carnahan, H. (2013). Self-directed practice schedule enhances learning of suturing skills. *Canadian Journal of Surgery, 56*, e142-e147.

Salmoni, A. W., Schmidt, R. L., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical appraisal. *Psychological Bulletin, 95*, 355-386.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological review, 82*, 225-260. doi:10.1037/h0076770

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning a behavioural emphasis* (4th ed.). Champaign, IL: Human Kinetics.

Schmidt, R. A., & Wrisberg, C. A. (2004). *Motor learning and performance*. Champaign, IL: Human Kinetics.

Shea, C. H., Wulf, G., & Whitacre, C. A. (1999). Enhancing training efficiency and effectiveness through the use of dyad training. *Journal of Motor Behavior, 31*, 119-125.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179-187.

Shea, J. B., & Titzer, R. C. (1993). The influence of reminder trials on contextual interference effects. *Journal of Motor Behavior, 25*, 264-274.

Shebilske, W. L., Regian, J. W., Arthur, W., & Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors, 34*, 369-374.

Shors, T. J., & Matzel, L. D. (1997). Long-term potentiation: What's learning got to do with it? *Behavioural and Brain Sciences, 20*, 597-655.

Summers, J. J., & Anson, J. G. (2009). Current status of the motor program: Revisited. *Human Movement Science, 28*, 566-577. doi:10.1016/j.humov.2009.01.002

Sweller, J., van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

van Kesteren, M. T., Rijpkema, M., Ruiter, D. J., Morris, R. G., & Fernandez, G. B. (1975). Building on prior knowledge: schema-dependent encoding processes relate to academic performance. *Journal Cognitive Neuroscience, 26*, 2250-2261.

Wagner, N., McQueen, S., & Sonnadara, R. (2018). Bridging the gap: theoretical principles behind surgical boot camps. In R. Rambani (Ed.), *Boot Camp Approaches to Surgical Training* (pp. 1-11). Basel: Switzerland: Springer.

Watkins, D. (1984). Students' perceptions of factors influencing tertiary learning. *Higher Education Research and Development, 3*, 33-50.

Woodworth, R. S. (1899). The accuracy of voluntary movement. *Psychological review, 3*, 1-119.

Wulf, G. (2007). *Attention and Motor Skill Learning*. Champaign: IL: Human Kinetics.

Wulf, G., Hob, M., & Prinz, W. (1998). Instructions for motor learning: Differential effects of internal versus external focus of attention. *Journal of Motor Behavior, 30*, 169-179.

Wulf, G., McNevin, N. H., Fuchs, T., Ritter, F., & Toole, T. (2000). Attentional focus in complex skill learning. *Research Quarterly for Exercise and Sport, 71*, 229-239.

Wulf, G., McNevin, N. H., Shea, C. H., & Wright, D. L. (2001). Learning phenomena: future challenges for dynamical systems approach to understanding the learning of complex motor skills. *International Journal of Sport Psychology, 30*, 531-557.

Chapter 3:

Using a Context, Input, Process, and Product Evaluation Model to Develop, Implement, and Evaluate a Boot Camp for Novice Surgical Trainees

3.0 Using a Context, Input, Process, and Product Evaluation Model to Develop, Implement, and Evaluate a Boot Camp for Novice Surgical Trainees

3.1 Introduction

As discussed in chapter one, the transition from medical school to residency is often considered the most difficult year for both teachers and learners in medical education. Across the world, learners report feeling underprepared and stressed when they take on their new role and responsibilities entering residency (Bligh, 2002; A. Cameron, Millar, Szmidt, Hanlon, & Cleland, 2014; Cave et al., 2009). Moreover, some studies suggest teaching hospitals experience increased patient mortality and decreased efficiency during this transition period (Young et al., 2011). Previous research suggests that boot camps (BCs) can be an effective way to ease the transition to residency (chapter 1.6). We also know that Surgical Foundations (SF) programs in Canada are uniquely positioned to implement curricular change (chapter 1.7.7). Taken together, we propose SF programs may be able to leverage the transition to competency-based medical education (CBME) frameworks to implement BC programs. However, before this can be done, there are a number of gaps in the BC literature that must be addressed.

Previous literature on BCs has focused on the short-term effectiveness of the programs through evaluating learner performance and learner experiences in BCs. There remains a paucity of literature on how BC programs were developed, the theoretical principles driving BC design, and the implementation processes (Blackmore, Austin, Lopushinsky, & Donnon, 2014). This lack of information limits our understanding of why certain aspects of BC programs are effective, while others are less so. It also limits the generalizability of findings, as educators are unable to implement the BC programs being studied into their own training contexts. The purpose of this study was to use an evidence-based approach to develop, implement and evaluate a SF BC program that adhered to the new Canadian CBME curriculum Competence By Design (CBD) discussed in chapter one, as well as the theoretical principles of novice skill acquisition discussed in chapter two. One way to capture the development and implementation processes is through systematic program evaluation, specifically the Context, Input, Process, and Product (CIPP) program evaluation model (Stufflebeam, 1971).

Formalized program evaluation is a fairly new research area (Stufflebeam & Shinkfield, 2007). Following World War II, federal agencies created a number of new national curriculum projects in science and mathematics. In order to help define the goals of these programs through needs assessments, examine the implementation and delivery of services, and capture the intended and unintended consequences of the programs, a number of new program evaluation models were created. Program evaluation models are sets of conceptual, hypothetical, pragmatic, and ethical principles, which form a framework to judge the overall 'worth' or 'value' of a program (Stufflebeam & Shinkfield, 2007). There are many program evaluation models available to researchers; however, this thesis focuses on the CIPP model for a number of reasons.

Firstly, the CIPP model can be used proactively to help develop a program, or retroactively to evaluate a program. As we wanted to both design and evaluate the BC program, having a model that could do both was extremely useful. Secondly, most program evaluation models do not account for contextual factors influencing a program and focus on providing a judgement of worth at the end of the evaluation process. The CIPP model on the other hand, was designed to evaluate educational programs occurring in dynamic social environments and dedicates an entire phase to understanding the context of a program (Gandomkar, 2018; Stufflebeam & Shinkfield, 2007). As medical training takes place in a unique environment where education is impacted by external factors in healthcare, it was important to capture the context. Lastly, the CIPP model aims to facilitate program improvement, rather than just provide a judgement of the program (Stufflebeam, 1971). In addition to this thesis being a research project, this work was part of a long-term initiative to improve surgical training; thus, using the findings to improve future iterations of the BC was critical.

The CIPP model is comprised of four phases: Context, Input, Process, and Product evaluation (Stufflebeam, 1971). **Context** evaluation assesses the needs, problems, and opportunities available to help delineate the goals and priorities of a program. The context phase determines "what needs to be done" within the given context (Frye & Hemmer, 2012; Stufflebeam & Shinkfield, 2007). **Input** evaluation assesses potential approaches, logistics, feasibility, and cost-effectiveness of a program. The input phase determines "how should it be done" within the given context (Frye & Hemmer, 2012; Stufflebeam & Shinkfield, 2007). **Process** evaluation assesses the implementation of a program and identifies areas for improvement. The process phase determines "whether it is being done as planned" (Frye & Hemmer, 2012; Stufflebeam & Shinkfield, 2007). Lastly, **Product** evaluation assesses the outcomes of the program to identify whether it was worth the investment. The product phase determines whether the program "succeeded" (Frye & Hemmer, 2012; Stufflebeam & Shinkfield, 2007). This chapter describes the development, implementation, and evaluation of a SF BC program, using the CIPP evaluation model.

3.2 Methods

3.2.1 Setting and Participants

This study involved residents from the McMaster Surgical Foundations (SF) program, in Hamilton, ON. The McMaster SF program is comprised of approximately 35 trainees in their first year of residency, and 35 trainees in their second year of residency, from nine different surgical specialties, including: Cardiac Surgery, Plastic Surgery, Obstetrics and Gynecology (OB/GYN), Vascular Surgery, Urology, Otolaryngology-Head and Neck Surgery (OTL-HNS), Neurosurgery, General Surgery, and Orthopaedics. Traditionally residents from the McMaster SF program participate in one academic half-day per week. Academic half-days are structured teaching sessions, which use a number of education strategies, including didactic, case-based, or lab-based/practical skills sessions (Wagoner & Seltz, 2019). During academic half-days the residents are free from their clinical responsibilities. Residents spend the remainder of the week completing clinical rotations at the direction of their home specialty. For residents at McMaster, those clinical rotations take place across four different hospital sites in Hamilton, ON. At the time of this study, the McMaster SF program was scheduled to transition to CBD in July 2017 (McMaster University, 2017).

3.2.2 Study Design

This study began in November 2015 and followed the CIPP program evaluation model, with four complementary phases. Throughout the four phases, data were collected from a variety of different sources, such as: the SF CBD Specialty Committee (described in chapter 1.7), McMaster Department of Surgery leadership, Program Directors and Program Coordinators from the McMaster Department of Surgery, faculty members from the McMaster Departments of Surgery, Medicine, and OB/GYN, and residents enrolled in the McMaster SF program in 2016. Methods such as focus groups, semi-structured interviews, surveys, and observational notes were used to address the research questions in the different evaluation phases. Table 1 provides a summary of the methods used in each phase, while detailed descriptions are provided below. Table 2 provides a summary of participant demographics. Figure 1 provides a timeline of study events. Development of the educational program was done in an iterative manner, thus the progression between phase one and two was not strictly linear. One of the benefits of using the CIPP model was that it facilitated this iterative process and allowed us to go back and forth between the context and input phases to identify the educational needs and gather feedback from the different stakeholders in a valid and reliable manner. This study was deemed exempt from requiring an ethics review, as the Hamilton Integrated Research Ethics Board (HiREB) considered it a quality improvement study.

3.2.3 Four Phases of the CIPP Evaluation Model

Context evaluation phase.

The goal of the context evaluation phase was to determine whether concerns about the transition from medical school to residency existed at a local level. To determine whether there were concerns, a meeting with the McMaster SF Program Director and SF Program Coordinator took place in November 2015. The SF Program Director is responsible for overseeing the education and evaluation of residents enrolled in the SF program. The SF Program Coordinator provides administrative support to the SF program. During the meeting, the SF Program Director and SF Coordinator shared their views on the transition to residency, the objectives of the SF program, educational needs, and anticipated challenges of transitioning to CBD. Following this initial meeting, a number of subsequent meetings took place with faculty and administrators from across the McMaster Department of Surgery (Figure 1).

In addition to these meetings, we approached current residents to discuss their views on the transition from medical school to residency. First-year residents enrolled in SF who were scheduled to complete an examination on June 1st, 2016, were invited to participate in a survey during their rest station (Appendix A). First- and second-year residents from SF who were not scheduled to complete the examination were invited to participate in a focus group occurring at the same time (Appendix B). Surveys and a focus group were selected as surveys provide participants with a place to offer their opinions anonymously, promoting candor, whereas focus groups can provide in-depth insight into participant perception, attitudes, and beliefs in an efficient manner; allow members to interact while sharing their experiences; and allow the researcher to explore themes that go beyond what was included on the focus group script (Khan et al., 1991). Including both methods gave us the opportunity to benefit from the strengths of both approaches and ensure we captured residents' true perceptions.

Participation was voluntary and did not impact the residents' standing or progression in SF. Verbal consent was obtained from all participants. Survey responses were anonymous and coded into Excel sheets. The focus group recording was transcribed verbatim and participant identifiers were removed. All qualitative data was then entered into NVivo (Version 12) for analyses. NVivo is a software program that can be used to sort many different formats of qualitative data (NVivo, 2019).

While there are many different approaches to qualitative research (Creswell, 2007), and considerable overlap among those approaches in terms of methods and analysis techniques (Vaismoradi, Turunen, & Bondas, 2013), qualitative description (QD) was used to analyze the qualitative data collected during the context phase of this study. Qualitative Description (QD) is not driven by pre-existing theories and does not try to generate a theory (Sandelowski, 2000). Rather, QD provides a rich, comprehensive summary of different perspectives on one event or experience based on the data collected (Neergaard, Olesen, Andersen, & Sondergaard, 2009). As the goal of the context phase was to explore faculty and resident perceptions on the transition from medical school to residency, and whether a BC would be useful, QD was selected as the most appropriate approach.

Input evaluation phase.

Determining how to best meet the educational needs of the SF program at McMaster University was done through an iterative process. We started with the principles of constructive alignment (Biggs, 1996). Constructive alignment is a theory in higher education, sometimes referred to as outcome-based learning or 'backwards' alignment, which focuses on aligning teaching strategies with intended learning outcomes. To achieve constructive alignment, we: 1) received input from the SF Program Director and Program Coordinator, Department of Surgery leadership, and the SF Committee on which skills should be included in the BC (i.e. what residents should know/be able to do at the end of BC); 2) determined how each skill would be evaluated (i.e. how to best capture whether they obtained that knowledge/could perform that skill); and 3) designed teaching sessions based on the assessment strategy selected and some of the principles of novice skill acquisition discussed in chapter two. These principles included: deliberate and distributed practice, observational learning, frequent formative assessments, and self-regulated feedback (chapter 2.4). Following this process, the draft schedule was presented to the SF committee in February 2016. The schedule was refined based on feedback from the SF committee and results of the resident survey and focus group in June 2016 (described in the context phase). The BC was implemented in July 2016. See Figure 1 for timeline.

Process evaluation phase.

The goal of the process evaluation phase was to determine whether the SF BC was implemented according to the plan. During implementation, three independent researchers took daily observational notes of each session. One of the observers was always the graduate student researcher (N.W.); the other two observers were undergraduate students, graduate students, or research assistants from our lab who volunteered in the BC based on their availability. All observers were provided with a list of prompts for what to look for during implementation - for example, did the session start on time? Were all scheduled facilitators present? The list of prompts can be found in Appendix C.

We also wished to confirm the trustworthiness of these observational notes. In qualitative research, trustworthiness is the accuracy, authenticity, or truthfulness of the data (Cypress, 2017). Trustworthiness is considered equivalent to the empiricist concepts of validity and reliability. Similar to validity and reliability, trustworthiness has multiple components. Components of trustworthiness include credibility, dependability, and transferability (Cypress, 2017). Credibility, or the accurate representation or truthfulness of data, can be established by comparing the original findings to secondary evidence as a way of triangulation (Cypress, 2017). We sought to establish credibility of the observational notes by comparing the data collected from the three independent observers for each session. We also reviewed themes from the observational notes with the BC observers during a debrief session on July 26, 2016 as a 'member-check' of the data. Dependability, or how stable the findings are over time, and transferability, or how generalizable the findings were to other contexts, were outside the scope of this CIPP study.

In addition to the observational notes, 'exit' interviews were conducted during the BC with session facilitators to discuss their experience in the BC session, as well as some discussion on the transition from medical school to residency more broadly. Facilitators were approached at the end of each session and asked to participate in a brief semi-structured interview (Appendix D). As the goal of the process phase was to use multiple perspectives to explore whether the BC was implemented according to the plan, qualitative description (QD) was once again selected as the most appropriate qualitative approach for both the observational notes and facilitator exit interviews.

Product evaluation phase.

After the BC was implemented, the product evaluation phase sought to determine whether residents who participated in the BC felt that the BC eased the transition between medical school and residency. To explore this, pre- and post-surveys were administered to the BC participants. The pre-survey was administered the first morning of the BC (Appendix E). The pre-survey included questions on participants' preparedness for residency, concerns entering residency, what they hoped to learn during the BC, as well as some additional questions on assessment practices in medical education and the transition to CBD. The post- survey was administered on the second-last day of the BC (Appendix F). The post-survey included questions on participants' preparedness for residency, whether the BC was a useful learning experience, what worked well in the BC, what could be improved upon, as well as questions on the assessment practices during the BC. Completing the survey was optional and all survey responses were anonymous. Participation did not influence residents' standing or progression in the SF program.

Qualitative description (QD) was used to identify main themes. Themes were then disseminated to the SF Program Director, SF Program Coordinator, Department of Surgery leadership and SF Committee to inform the future iterations of the program. Additional work exploring the longer-term effectiveness of the BC will be presented in chapter four.

3.2.4 Analyses

Qualitative data.

There are several ways to analyze data when using QD. Most commonly, QD studies cite using either thematic or content analysis. Within the qualitative methodological literature, these are sometimes used synonymously (Sandelowski & Leeman, 2012), whereas other authors have attempted to define and differentiate the two (Braun & Clarke, 2006; Hsieh & Shannon, 2005). For this study, conventional content analysis was used to analyze the qualitative data collected in the context, process, and evaluation phases.

Conventional content analysis is a subtype of content analysis, which is generally used with study designs aiming to describe an event or experience from multiple viewpoints. Furthermore, this approach is appropriate when existing theories explaining the event or experience are limited. Using conventional content analysis, researchers identify categories based on the data, rather than creating categories based off prior knowledge (Hsieh & Shannon, 2005). Conventional content analysis works best when questions and probes are open-ended - for example, asking participants "Can you tell me more about that?" during an interview or focus group. As the main purpose of this study was to describe the development and implementation of a SF BC using the CIPP evaluation model, and since little information exists on these processes to date, the conventional content analysis method was deemed appropriate for all evaluation phases.

As described by Hsieh and Shannon (2005), conventional content analysis starts with the researcher reading all of the data repeatedly to gain an understanding of the whole picture. Once that has been achieved, the researcher organizes the data through open coding. Open coding is the process of writing notes or highlighting specific words while reading the data, and then using those notes or words to generate categories. These categories are grouped into themes based on how they are related or linked. Themes are used to create a general description of the event or experience, using language that is relatable and true to the study participants' perspectives. The graduate student researcher (N.W.) completed this process for all qualitative data.

Quantitative data.

To address whether BC participants' self-rated preparedness differed on the pre- versus post-BC surveys (phase 4), descriptive statistics were calculated. As the surveys were anonymous in order to promote candor, they could not be paired. Furthermore, as participants ranked their preparedness on a Likert scale, the data were analyzed using a Mann-Whitney U test. The analysis was completed using SPSS Version 25. Statistical significance was considered at $p \leq .05$.

3.3 Results

3.3.1 Context Evaluation Phase

Faculty perspectives.

During the meeting with the McMaster SF Program Director and Program Coordinator in November 2015 (see Figure 1 for timeline), they suggested that surgical residents at McMaster could be better supported during their transition from medical school to residency. They also acknowledged that changes needed to be made to the SF curriculum as the program was set to transition to CBD in July 2017. Based on previous research (see chapter 1.6), they were in agreement that a BC for novice surgical trainees could achieve both of those goals.

Following the initial meeting with the SF Program Director in November 2015, McMaster Department of Surgery leadership was contacted for their input and approval of a BC for incoming surgical trainees. The Department of Surgery leadership agreed with the need for a BC and approved the project. Next, a meeting was conducted with the SF committee. The SF committee is comprised of Program Directors and Program Coordinators from all surgical specialties involved in the SF program. The SF committee oversees any changes to the SF curriculum; thus their input and approval was required. Once again, the SF committee agreed with the need for a BC and approval was granted.

Resident perspectives.

Demographics.

Of the 16 first-year residents completing the examination on June 1, 2016, 16 residents (100%) participated in the survey. Twenty-two residents were assisting with the examination and were unable to participate in the survey or focus group. Of the remaining 32 SF residents, 22 were absent, however the other 10 agreed to participate in the focus group (Table 2). The 10 focus group participants included five first-year residents and five second-year residents. Together, the survey and focus group represented 37% of the SF program. Specifically, as 21/26 participants were in their first year of training, our sample represented 60% of the first-year residents in SF at McMaster. Themes and sample quotations are presented in Table 3.

Varied preparedness.

Survey participants ranked how well they thought medical school prepared them for residency on a scale of 1 (not prepared at all) to 10 (completely prepared). Survey participants rated their preparedness for residency as fairly high (M = 7.06; Mode = 8). However, responses ranged from 3/10 to 9/10. Focus group participants reported similar levels of perceived preparedness for residency, suggesting that medical school prepared them “somewhat” for residency, but feelings of preparedness varied between individuals. Participants suggested this was due to individuals entering residency with different clinical experiences. Within the same medical school, learners have different experiences due to the unpredictable nature of what comes into the hospital or clinic on any given day. Between medical schools, learners have different experiences due to the size of the program/number of other learners, exposure to different rotations/specialties, and different expectations about the role of medical students in the clinical environment (quote 1.1.1). Focus group participants suggested that all of these factors influence how prepared someone feels for residency.

Areas needing additional support at the onset of residency.

Two of the sixteen survey participants reported they felt they were adequately supported during the transition from medical school to residency. However, the remaining 14 survey participants and all 10 focus group participants identified several areas where they wished they had additional support when starting residency.

Knowing who to call for help (and when).

Focus group participants said that as a new resident, they did not necessarily know what resources and services were available at the hospital. For this reason, participants said that they often felt alone/lost during their first few call shifts. This was particularly true for learners who completed their medical school training at a different University. Residents said that knowing there were services available to help, and when to call those services, would have been a huge source of support for them at the onset of residency (quote 2.1.2).

Understanding hospital logistics.

Participants reported that having additional information on general hospital logistics would have been extremely helpful for incoming residents. Hospital logistics included information on the layout of the hospital, electronic medical record systems, and the order and admission forms. The need for support in this area seemed particularly relevant for residents who had completed medical school in another province or country (quote 2.2.2).

Refresher on clinical knowledge.

In addition to not knowing about hospital-specific factors, participants reported that their clinical knowledge and skills were 'rusty' at the start of residency. After having approximately two months off between medical school and residency (sometimes more), participants suggested it 'took a while to remember things'. Participants suggested reviewing basic clinical knowledge/skills in a BC type program could get everyone on the same page, allow residents to contribute to patient care in a safe manner, and accelerate future clinical learning as less time would need to be spent reviewing the basics (quote 2.3.2).

Implementing clinical knowledge/skills in practice.

However, participants noted that having clinical knowledge/skills and implementing clinical knowledge/skills are two different things. It is difficult for new residents to manage situations that are not 'textbook' or manage situations when there are multiple things happening at once. As a medical student they felt somewhat sheltered from this challenge (quote 2.4.1).

3.3.2 Input Evaluation Phase

To determine the best approach to meeting the learning objectives, the principles of construct, or 'backwards' alignment were followed. This included: 1) determining the BC intended outcomes; 2) determining the BC assessment strategies; and 3) designing the BC teaching strategies.

Determining intended boot camp outcomes.

To determine what learners should know/be able to do by the end of the SF BC, input was received from a variety of stakeholders. The SF committee suggested the BC address the SF Transition to Discipline Entrustable Professional Activities (EPAs), which are tasks that surgical residents must be able to do within their first stage of training under the new CBD mandate (see chapter 1.7). The SF Transition to Discipline EPAs included skills such as: pre-operative preparation; recognition and early management of a critically ill patient; documentation of clinical encounters; handover; ability to function in the operating room; wound closure; management of tubes; drains and lines; and post-operative care.

In addition to the Transition to Discipline EPAs, the SF Committee wished to include other skills they believed were critical for trainees to know at the onset of residency. These skills included breaking bad news, informed consent, critical incident management, electrosurgery, surgical instrumentation, radiation safety, and the Advanced Trauma and Life Support (ATLS) course.

During the June 2016 survey and focus group (described above), residents agreed that including sessions on resuscitation, chest tubes, lines, surgical instrumentation, pre-operative preparation, informed consent, and post-operative care would be useful at the onset of residency. Residents also suggested that sessions on hospital logistics (electronic medical record systems, order and admission forms, resources/consulting services available), ward management, how to teach/supervise medical students, specialty-specific skills, and a resident-led ‘what I wish I had known’ session would be helpful. Figure 2 outlines the topics suggested by each stakeholder group and identifies where their recommendations overlapped.

Determining assessment strategies.

In terms of how learners were to be assessed during the BC, the SF committee agreed that frequent, formative assessment should be included in each BC session. Using formative assessment would adhere to the new CBD guidelines on assessment (see chapter 1.7.5), as well as best practices in the novice skill acquisition literature (see chapter 2.3.6).

In addition to formative assessments during the BC sessions, the SF committee decided that an Objective Structured Clinical Examination (OSCE) would be included at the end of the two-week BC. Objective Structured Clinical Examinations (OSCEs) are an evaluation technique where learners rotate through a series of simulated scenarios (or stations) to demonstrate their clinical skills (Harden, Stevenson, Wilson Downie, & Wilson, 1975), and are currently considered the gold standard for evaluating clinical skills in medical training. Nine OSCE stations were developed based on the BC sessions, including informed consent, pre-operative preparation, suturing, breaking bad news, operating room conduct, recognition of the critically ill patient, handover, a multiple-choice quiz on surgical instrumentation, electrosurgery, and radiation safety, and a rest station. Most of the stations (5/9) targeted the SF Transition to Discipline EPAs outlined in Figure 2. During OSCEs, resident performance is typically measured using a modified Objective Structured Assessment of Technical Skills (OSATS) (Appendix G). The OSATS was developed by Martin et al. (1997) and includes a dichotomous task-specific checklist as well as several Likert scales. Chapters four and five will discuss OSCEs and the OSATS in more detail.

Designing teaching sessions.

For the BC sessions that were associated with OSCE stations, items from the task-specific checklists were turned into session objectives. For example, one item on the informed consent OSCE station checklist was "correctly completed a Hamilton Health Sciences consent form". This became one of the informed consent session objectives (Appendix G). Once the objectives of each session were outlined, the sessions were structured to include the principles of novice skill acquisition discussed in chapter two. To facilitate deliberate practice (chapter 2.3.1), the goals and objectives of each session were to be posted during the sessions. Facilitators were meant to direct the residents back to the objectives, and (when appropriate) challenge the residents with more complex tasks. In terms of practice schedules (chapter 2.3.3), the BC was designed so practice was distributed, rather than massed. For technical sessions this meant residents were scheduled to rotate through different tasks during the session. For example, in the tubes and drains session, residents were to spend a set period of time on chest tubes, airways, and central lines, prior to an open practice session. For non-technical skills sessions, residents were supposed to rotate through different case-based scenarios. The hope was that distributed practice would promote encoding and retrieval through contextual interference (chapter 2.3.4). Throughout the BC residents were also supposed to work in pairs, observe each other, and provide feedback. When the pair felt ready, they could demonstrate their skills to a facilitator. This approach was meant to promote observational learning (chapter 2.3.2), and self-directed feedback (chapter 2.3.7). Figure 3 depicts the finalized 2016 BC schedule. The BC was implemented in July 2016.

3.3.3 Process Evaluation Phase

To determine whether the BC was implemented according to the plan, 26 observational notes and nine facilitator exit interviews were collected and analyzed using conventional content analysis. Themes and sample excerpts are presented in Table 4.

Daily observations.

Based on the daily observations collected, the BC was, for the most part, implemented according to the plan. All sessions started on time, though many sessions did not take as long as expected (excerpt 3.2). With the exception of one facilitator who was called to an emergency surgery (pre-operative preparation – surgery didactic), all facilitators were present for their scheduled sessions. Observers noted that most sessions followed the intended plan; however there were often one or two minor objectives missed (excerpt 4.1). In terms of resident engagement, residents seemed particularly attentive during resident-led sessions and sessions including case-based examples; however, resident engagement tended to decrease by the end of each session regardless of the topic or facilitator (excerpt 6.2). Although not formally part of the curriculum, observers also noted that residents used the BC sessions as an opportunity to socialize and get to know one another (excerpt 7.1). Lastly, observers made recommendations for future iterations of the BC. Observer recommendations often focused on logistics (room, equipment set-up), case variability (e.g. including pediatrics and obstetrics cases), making sessions more interactive, and adjusting the length of sessions (excerpt 8.2).

Facilitator exit interviews.

The BC included 61 facilitators from the Departments of Surgery, Medicine, and OB/GYN. These facilitators included two fellows, four allied health professionals, 23 senior residents, and 32 attending physicians. When facilitators were approached about the exit interviews, they often stated that they had to get back to their clinical responsibilities and thus did not have time to participate. However, nine facilitators (15%) agreed to participate in an exit interview immediately following their BC session. All facilitators had volunteered their time. Table 5 provides a summary of themes and sample quotations.

In terms of the transition between medical school and residency, 100% of facilitators said they notice a gap for some students, but not others. Facilitators believed this was due to the varied experiences learners have as medical students (quotation 9.1.1). Facilitators suggested that a BC alone might not be able to close this gap, as there are some skills that just 'take time to learn' (quotation 9.2.2). That being said, 100% of facilitators thought the BC was a worthwhile program and were willing to come back to participate in future iterations.

3.3.4 Product Evaluation Phase

To explore whether the BC eased the transition from medical school to residency, 32 BC participants (100%) completed the pre-BC survey and 31 (97%) completed the post-BC survey. Table 6 provides a summary of themes and sample quotations.

Preparedness self-rating.

As previously mentioned, the surveys were anonymous, so the results could not be paired. However, the results of the Mann Whitney U test suggested that participants' self-rated preparedness was significantly different between the pre-BC survey (Mdn = 6), and the post-BC survey (Mdn= 7, U= 302, $p < .01$). Figure 4 provides an illustration of the preparedness results.

Resident perspectives on the boot camp.

Of the 31 residents who completed the post-BC survey, 28 (90%) reported the BC was definitely useful for incoming residents, while 3 (10%) said the BC was somewhat useful (quotation 11.1.2). Residents felt the structure of the BC, with didactics in the morning and lab sessions in the afternoon, worked well. Moreover, residents liked the sessions that included practical 'hands-on' information, small group breakouts, and mock patients. Residents identified the breaking bad news, informed consent, handover, and recognition of the critically ill patient sessions as particularly helpful (quotation 12.31). In terms of what could be improved, residents suggested the timing of sessions could be adjusted as some sessions were too long and some were too short. Furthermore, residents perceived the BC as being very general surgery focused (quotation 13.2.1). Residents suggested the BC content could include examples from different specialties to make it feel applicable for everyone. In addition to this, residents suggested the Advanced Trauma and Life Support (ATLS) course be moved from the weekend to during the week, and some suggested the overlap between ATLS and the tubes and drains session be removed. Residents also suggested the operating room conduct, electrosurgery, and pre-operative preparation sessions could be modified/shortened to accommodate the inclusion of ATLS. Lastly, when asked what (if any) concerns residents still had regarding residency, residents reported still feeling concerned about time management in the clinical environment, caring for sick patients independently, navigating the hospital systems, and specialty specific knowledge/skills (quotation 14.2.2).

3.4 Discussion

3.4.1 Value of the CIPP Model

When examining new educational programs, researchers often use pre-post performance or perceptions of a program to determine whether or not ‘it worked’. However, relying on these outcomes in isolation can miss valuable information on the context, development, and implementation of the program, all of which can significantly impact program effectiveness (Frye & Hemmer, 2012). In this study, we drew upon the program evaluation methods, specifically the CIPP program evaluation model, to explore the development, implementation, and outcomes of a BC for incoming surgical trainees. Through comparing the results of the evaluation phases, a number of key findings emerged.

Feelings of preparedness vary between residents.

In this study, two cohorts of residents rated their perceived preparedness for residency. During the context phase, 16 SF residents rated their perceived preparedness for residency at the *end* of their first year. During the product phase, 31 incoming SF residents rated their perceived preparedness on the *first day* of residency. Interestingly, residents at the end of their first year rated their perceived preparedness for residency as higher than those on their first day of residency. While it is possible that this is a result of differences amongst the cohorts, it is also likely that residents at the end of residency experienced some recall bias and had forgotten how worried they felt when starting residency. If recall bias influences learners' perceptions of this transition after only one year, it is not surprising that some faculty dismiss the need for a transitionary program for incoming residents. It has been quite some time since faculty had their first day of residency, and they may not accurately remember exactly how they felt. This may be one of the barriers preventing programs from successfully implementing BCs for incoming trainees. Strategies for overcoming barriers and implementing successful curricular change are discussed in more detail below.

In addition to the variability in perceived preparedness between the two cohorts, there was a wide range of preparedness scores within each cohort. Residents suggested this was likely because perceived preparedness is a highly individual thing. Residents thought the reason for these individual differences were, at least in part, due to medical students coming to residency with different clinical experiences.

Medical students graduate with different experiences.

Throughout this study, the idea that medical students graduate with different experiences came up multiple times. This theme was present in the surveys and focus groups with residents in the context phase, along with facilitator exit interviews in the process phase. All participants suggested that experiences differ both within and between medical schools. Within medical schools, clinical encounters are unpredictable, thus students will see different patients with different complaints. Little can be done to mitigate this. However, between medical schools, varied experiences seem to be due to the lack of standardized medical school curricula. Some medical schools in Canada are different lengths (e.g. three or four years); include different clinical rotations; use different examination processes; and have different expectations on the role of medical students in the clinical environment. There is no national curriculum, particularly for the clinical portion of medical schools. Thinking back to the history of medical education discussed in chapter 1.2, this lack of standardization is not a new issue. The Flexner Report highlighted a number of areas where medical schools were not standardized (Flexner, 1910). Thus, while the Flexner Report resulted in a lot of changes, such as the inclusion of basic sciences and clinical rotations in the medical school curriculum and national medical licensing examinations, it seems we have still not achieved a standardized curriculum.

The boot camp provides a baseline for residents.

Due to the varied experiences in medical school, it was suggested that one of the major benefits of a BC is providing residents with a ‘baseline’. Participants felt the BC was a way to ‘make sure everyone was on the same page’. Results from both the process and product phases suggested that this was particularly true for the sessions that provided practical, or ‘hands-on’ information. For example, residents identified the breaking bad news, informed consent, suturing, recognition of the critically ill patient, and handover sessions as particularly useful in establishing a minimum level of competency. This helped learners feel more confident and safer to provide patient care. Reflecting on the structure of those sessions and the daily observational notes, those sessions closely followed the principles of novice skill acquisition discussed in chapter two. For example, all of the aforementioned sessions included a didactic component, along with a breakout session, where residents where split into small groups or pairs and worked through a number of cases or skills (distributed learning). During these breakout sessions, residents were encouraged to observe their peers (observational learning), focus on specific aspects of the skill at hand (deliberate practice), and ask facilitators for feedback when they felt ready (self-directed feedback). From this, we suggest that following the principles of novice skill acquisition was critical to success. That being said, we also uncovered evidence suggesting that BCs may not be able to address all skills.

Some things take time.

Within our study, participants often referred to certain skills 'taking time' or 'experience' to learn. During the context phase, residents suggested that having medical knowledge and implementing that knowledge are two very different things. Residents suggested that implementing their knowledge in the clinical environment was one of the biggest challenges they faced during the transition to residency. Thus, while residents thought the BC was helpful and might accelerate the process, figuring out how to implement their knowledge will still take time. During the process phase, facilitators echoed this sentiment, as they suggested BCs could provide learners with some basic knowledge, however a BC will never be able to completely prepare learners for residency. To become truly comfortable, facilitators suggested it takes time and experience in the clinical environment. The post-BC surveys reflected this idea as well. While resident self-rated preparedness scores increased when compared to the pre-BC survey, there were no residents who scored themselves as 10/10 prepared for residency following the BC. Furthermore, within the written comments, residents identified a number of areas where they still had concerns, including time management, care of critically ill patients, and making decisions independently. Thus, while BCs are not able to fully prepare learners for the transition to residency, BCs can provide residents with basic knowledge, tips, resources, and coping strategies for things they will encounter in the clinical environment. It is also important to note that feeling completely prepared or confident is not always a good thing for junior trainees. Having an overconfident trainee can be just as dangerous, if not more so, than an under confident trainee. The BC seemed to provide trainees with some insight on their skill level, reaffirm what they are expected to be able to deal with, and clarify who and when to call for help.

Social component of boot camps.

In addition to the BC providing residents with the opportunity to acquire basic skills and information, the social aspect of the BC was often mentioned as a major benefit. Observers noted that the BC, particularly the small group sessions and peer observations, facilitated resident interactions. During the post-BC surveys, residents reaffirmed those observations, commenting on how they were extremely grateful for the opportunity to meet their co-residents, other senior residents, and faculty in a low-stress environment. In medical training, we know that individuals enter residency programs with different backgrounds and experiences, from different schools, and sometimes different countries. We also know that burnout rates are extremely high amongst junior physicians (chapter 1.4). Thus, while it was not a formal objective, we conclude that an extremely important outcome of the BC program was the opportunity for learners to create a social support network at the onset of residency. Having this social support network has the potential to improve residents' well-being, and also improve their clinical care, as they know who to talk to and feel comfortable initiating those conversations.

3.4.2 Organizational Socialization

Bringing these overarching themes together, it seems that individuals have varying levels of perceived preparedness, and perceived preparedness seems related to experiences in medical school. Thus due, in part, to the variability within and between medical schools, some individuals experience more difficulty than others when transitioning to residency. Boot camps seem to ensure residents have a base level of information and skills and help 'level the playing field'. This is particularly true if the BC sessions focus on delivering practical information through the principles of novice skill acquisition. In addition to this, BCs seem to be extremely useful in creating new social networks. One theory that highlights why these things are important for new residents is Organizational Socialization.

Organizational Socialization comes from the social psychology literature and refers to the process of learning the behaviours, attitudes, and social etiquette needed to assume a role within an established organization (Van Maanen & Schein, 1979). Four components of Organizational Socialization include: task mastery, role clarification, acculturation, and social integration (Morrison, 1993). Addressing some of the basic skills/information in a BC setting will likely allow residents to focus on more advanced skills in the clinical environment (i.e. task mastery). The resident-led sessions, which were preferred among BC participants, helped explain what is expected of new residents (role clarification) and also provided 'insider information' into how residency works (acculturation). Lastly, the BC provided the environment for incoming residents to meet their co-residents and faculty in a low-stress environment, promoting the development of new relationships (social integration).

Morrison (1993) states that these components of Organizational Socialization are extremely important for anyone starting a new role, but is critical for those individuals working in highly-structured environments with their own culture and rules. As healthcare exists within a unique culture and hierarchy, we propose including Organizational Socialization for newcomers at any stage in healthcare is important. Yet, we rarely see formal Organizational Socialization in healthcare. This might be part of the reason why teachers and learners suggest that transitions in medical training are extremely challenging (chapter 1.5). For those transitioning into residency, our results suggest that Organizational Socialization in the form of a BC has the potential to improve resident comfort levels, patient care, and ensure residents can get the most out of their clinical learning experiences. Sharing the results of this study, and framing them within Organizational Socialization may provide the impetus for long-term curricular change during the transition to residency. Beyond that, these results may inspire Organizational Socialization at other transition points in medical training.

3.4.3 Curricular Change in Medical Education

However, it is important to recognize that long-term curricular change is not easy. It has been said that, "It is easier to move a graveyard than change a curriculum" (Bland et al., 2000, p. 592). Within the context of medical training this is particularly true, as curricular change is inhibited by the conflict between education and service provision (chapter 1.4). Moreover, the medical field is fixated on evidence-based practice. Whether it relates to clinical practice or teaching strategies, medical practitioners want evidence before implementing change (van der Vleuten, Dolmans, & Scherpbier, 2009). For clinical practice, there are clinical trials to collect this evidence. However, for educational interventions, there is no equivalent to randomized controlled trials, and it often takes a long time (and many people) to gather information. This has resulted in a stalemate when it comes to curricular change in medical education. Change is disruptive; thus educators want evidence to know it works before implementing the change. However, to gather evidence we need to implement change.

In our study, we were able to leverage the shift to CBD in Canada to enact change at McMaster that might not otherwise have been possible. Even so, this was really only possible because there was some previous evidence to suggest implementing a BC program may ease the transition from medical school to residency (chapter 1.6). Through using the CIPP program evaluation model, our study provides educators with evidence on the need for a BC type program at the onset of residency, as well as a detailed description of the development and implementation processes. We hope that through sharing this information, other educators will be able to promote change at their own institutions, and modify the BC program to their own training context. To be successful in this endeavor, there are a number of factors of which educators must be aware.

Bland et al. (2000) conducted a literature review and identified 35 features of successful curricular change within medical training contexts. Through using the CIPP model, we were able to describe how many of these features were addressed in our BC program (Table 7). However, there are a few take-home points. First, many of the 35 features concern organizational factors. This highlights again how important the context piece is to program evaluation, and the usefulness of taking a CIPP approach to program evaluation. Within the organizational factors, Bland et al. (2000) describe leadership, politics, and the involvement of members from different departments as being critical to success. This BC project was fortunate to have significant support from the Department of Surgery. This was, in part, due to the upcoming transition to CBD, but also because the need to support learners through the transition to residency resonated with many people. All facilitators took time away from their clinical practice to volunteer for this initiative. Because of this, there was the feeling of a shared mission or goal. If educators wish to implement a BC type program, ensuring they have organizational support is key.

In addition to organization factors, Bland et al. (2000) discuss how curricular change is a dynamic and complex process and that using an evaluation method is critical to success. Through using a CIPP method, we were able to provide educators with a comprehensive overview of the BC program, including the benefits, challenges, and areas for improvement. This provided everyone with the opportunity to give feedback on the initiative and generated the evidence for the Department of Surgery to make a decision on long-term implementation of the program. Thus, giving individuals the opportunity to voice their concerns and collecting evidence to inform decisions is key in curricular change. However, Bland et al. (2000) go on to state that enacting curricular change at the institution level often comes with unique and unanticipated challenges, and usually takes longer than expected. Educators must be mindful of the long-term outcomes and implications of their program. This study provides a snapshot of the 2016 SF BC; information is still needed on the long-term impacts. Long-term outcomes will be the focus of chapter four.

3.5 Limitations

This study had a number of limitations. First, as we had to ensure sufficient buy-in from Department of Surgery leadership and the SF committee to make this project a reality, BC planning started before the survey and focus group with current residents. For this reason, a lot of the BC schedule was already in place, and some of the resident suggestions such as 'how to teach medical students' had to be omitted from the 2016 BC schedule. Second, the BC facilitators were volunteers (as is customary for the SF curriculum), with varying degrees of experience in educating residents. While we feel that having volunteer facilitators increased buy-in for the BC program and gave our study high ecological validity, we were unable to control for factors such as presentation quality and adherence to the learning objectives, which may have influenced our data. Third, all four phases of this study had multiple components, and to immerse oneself in the data was extremely time-consuming and onerous, and thus only one researcher (N.W.) coded the qualitative data. While only having one coder is a potential threat to the trustworthiness of the data, other strategies such as member-checking, triangulation, and presentation of direct quotations (Tables 3-6) were used to minimize the risk. Fourth, resident self-rated preparedness scores were anonymous; while this anonymity promoted honest responses, we were unable to run paired statistical analyses. As our data suggest perceived preparedness is highly individual, paired data should be explored in future studies. Lastly, in light of the context specificity of the CIPP model, this study did not explore the transferability of the results. Thus, the BC may not function the same way in another program.

3.6 Conclusions

In summary, the previous literature suggests trainees need to be better supported during the transition from medical school to residency. While literature suggests BCs have the potential to provide such support, there has been little literature exploring how BC programs are developed, the theoretical principles driving BC design, and the implementation processes. This study was the first of its kind to use a CIPP program evaluation model to describe the development, implementation and short-term outcomes of a BC for novice surgical trainees. Through the Context evaluation phase, we found both faculty and residents at McMaster agreed with the need to better support trainees through the transition from medical school to residency. During the Input evaluation phase, we used constructive alignment to identify program outcomes, identify assessment strategies, and design teaching sessions based on the input from faculty, residents, and principles of novice skill acquisition discussed in chapter two. In the Process evaluation phase, daily observational notes and exit interviews with facilitators suggested the BC was implemented, for the most part, according to the intended plan and that adherence to pedagogical principles was critical. During the Product evaluation phase, we identified that the BC was largely successful as it significantly improved residents' self-rated preparedness for residency, and that the majority of residents (90%) felt the BC was extremely useful to their education.

Using the CIPP approach, we were also able to identify a number of overarching themes across the four phases. These included the idea that perceptions of preparedness for residency are highly individual. Medical students graduate with varied experiences. Providing residents with basic knowledge and information in a BC setting improves resident confidence and has the potential to improve patient care and accelerate later clinical learning as residents can spend cognitive resources on more complex tasks. This is especially true if BC sessions focus on practical information and adhere to the principles of novice skill acquisition. Lastly, the social aspect of BCs is critical in creating support networks and facilitating these benefits.

We hope that through sharing this information, educators are provided with the evidence to promote curricular change in their own programs. Moreover, through using the CIPP model to provide a detailed description of the BC program from start to finish, we hope educators will be able to modify the program to their own training contexts and we can continue to explore best approaches for BCs for novice learners, and move the field together as a whole.

3.7 Epilogue

Following the 2016 BC, major findings were disseminated to the SF Program Director, Program Coordinator, Committee, and Department of Surgery leadership. As the response to the BC was overwhelmingly positive, it was determined that BC would be updated based on the feedback, and re-run with the next incoming cohort in July 2017. Major findings and modifications are listed in Table 8. The 2017 SF BC was also used to explore different assessment strategies and how they relate to OSCE performance (discussed in chapter five).

During the implementation of the 2017 BC, pre- and post-boot camp surveys were once again completed; however, this time residents were asked to include their names. This allowed us to pair the data before it was de-identified and address one of the limitations with the 2016 BC data collection. The results of the Wilcoxon ranked sum test suggested that once again there was a significant difference in self-rated preparedness scores before ($M = 5.57$) and after the boot camp ($M = 6.65$); $z = 4.45$; $p < .01$.

Once the surveys were de-identified, the qualitative responses were also analyzed using conventional content analysis. Following the analysis, the results were disseminated to the Department of Surgery faculty. As the response to the 2017 BC and buy-in from the different programs that was even more positive than in 2016, the Department of Surgery decided the BC, with a few more modifications (Table 9), would be formally adopted into the SF training program for all incoming cohorts. A team was put in place to take over from the graduate student researcher (N.W.), and the third iteration of the boot camp was implemented in July 2018.

Table 1

Summary of Methods

Phase	Phase 1: Context	Phase 2: Input	Phase 3: Process	Phase 4: Product
Evaluation Question(s)	What are the educational needs of the SF program at McMaster University?	What is the best approach to meeting the educational needs of the SF program at McMaster University?	How was the educational intervention (boot camp) implemented compared to the plan?	Was the educational intervention (boot camp) successful? What changes (if any) need to be made to the boot camp?
Methods	1. Review of relevant literature 2. Expert panels 3. Resident survey 4. Resident focus group	1. Review of relevant literature 2. Expert panels 3. Resident survey 4. Resident focus group	1. Daily observation by 3 independent researchers 2. Facilitator Interviews	1. Pre- boot camp resident surveys 2. Post- boot camp resident surveys

Table 2

Participant Demographics

CIPP Phase	**Data Collection**	**Number of Participants**	**Gender (% Female)**	**Gender (% Male)**
Context & Input	Expert Panel	26/26 (100%)	12/26 (46%)	14/26 (54%)
	Resident survey	16/16 (100%)	8/16 (50%)	8/16 (50%)
	Resident focus group	10/32 (31%)	2/10 (20%)	8/10 (80%)
Process	Facilitator interviews	9/61 (15%)	8/9 (89%)	1/9 (11%)
Product	Pre-boot camp survey	32/32 (100%)	14/32 (44%)	18/32 (56%)
	Post-boot camp survey	31/32 (97%)	Not available	Not available

Note: The pre- and post-boot camp surveys were anonymous. While we can infer the gender proportions in pre-survey, we do not know which participant did not complete the post-survey, thus cannot comment on the gender proportions.

Table 3

Context Evaluation Phase: Resident Perspectives on the Transition to Residency

Theme	Sub-theme	Sample Quotations
Preparedness for residency (1)	Perceived variability in preparedness (1.1)	*"…I think the biggest thing is the day-to-day ward management stuff. Depending [on] where you went [to medical school], and what kind of rotations [you had], you may or may not have been involved in those clinical activities..." – Resident focus group (1.1.1)* *"I think it depends on where you go, but even within medical schools now they are starting to have such different campuses, like satellite campuses and things like that. It really becomes a huge variation in how prepared you are." – Resident focus group (1.1.2)*
Areas needing additional support at the onset of residency (2)	Knowing who to call for help and when (2.1)	*"Not graduating from McMaster… [someone said] go consult 'Thrombo'… I had no idea who they were talking about." – Resident focus group (2.1.1)* *"Often we felt like we were completely responsible for dealing [with] or treating a crashing patient …to understand that we have resources that may help us in dealing with difficult situations if you feel like you are alone in the middle of the night…is really, really great for an incoming resident." – Resident focus group (2.1.2)* *"The other thing is when to call for help. Most of the residents are not sure [of something]…but they don't want to call the staff or their senior. It seems to be simple, but I think that it is very important." – Resident focus group (2.1.3)*
	Understand-ing hospital logistics (2.2)	*"Would have liked an intro to the Ontario Medical system, orders, admissions, computer system, how hospitals work, entrapment area, appropriate consults, hospital layout, list of numbers [for example] paging, etc...." – Resident Survey (2.2.1)* *"As an International Medical Graduate, the Canadian system is new to me. It took me time to fully adapt." – Resident focus group (2.2.2)*
	Refresher on clinical knowledge and skills (2.3)	*"Reviewing the things that are expected of you as a surgical resident, you know going into the OR, that if you really mess up [sterility] it can potentially hurt the patient, or the OR time itself…. being able to understand what you are supposed to do in the surgical field is key." – Resident focus group (2.3.1)* *"It would be nice to be there on the first day and actually be able to help and be part of the team. And they are really basic things [to review] and it wouldn't take a lot of organizing." – Resident focus group (2.3.2)*
	Implement-ing clinical knowledge and skills in practice (2.4)	*"The problem is when you start residency, you don't practice what the textbook has taught you. [The] textbook and residency are two different things…when you are running around it is not the same way at all." — Resident focus group (2.4.1)* *"What they need to learn in a boot camp isn't the textbook stuff. It's the basics of what they need to know going into a critical situation on the floor in terms of management. … how to manage that situation as a resident, as a doctor." – Resident focus group (2.4.2)*

Table 4

Process Evaluation Phase: Daily Observations

Theme	Sample Excerpts
Session timing (3)	*"The length of the session was too long. By the end of the presentation, many residents were on their phones or lost attention." – Electrosurgery didactic (3.1)* *"The residents finished the evaluation forms at 3:00 PM [1 hour early]. Some residents stayed behind for [an informal] talk on Citrix while others with experience with the [software] left early." – Tubes and drains lab (3.2)*
Whether sessions followed the intended plan (4)	*"Most objectives were met, but there was no mention of respiratory therapists or speech and language pathologists even though they were on the paper schedule." – Ward management (4.1)* *"Deviated from the objectives a bit." – Operating room conduct (4.2)*
Resident Engagement (5)	*"Interspersing [the] presentation with content-specific questions and personal experiences in the field was effective in maintaining attentiveness among the residents." – Pre-operative preparation (5.1)* *"Most residents seemed attentive at the start. A few residents were texting at the beginning of the session, but re-engaged during the interactive components…By the end, the audience looked a little tired; a few people were sleeping and another few were looking down or texting." – Handover (5.2)*
Resident-led sessions (6)	*"Having Dr. XX share stories about his own mistakes as a resident and advice during different procedure was really useful and receptive among the residents." – Pre-operative preparation (6.1)* *"I have noticed over the past few days that the PGY1s are receptive of the PGY2s and ask all kinds of questions. Thus groups that were facilitated by an upper-year resident were more talkative and slower than groups facilitated by surgeons." – Electrosurgery lab (6.2)*
Social aspects of the boot camp (7)	*"Once tables had finished with their forms, the conversations got off-topic, but the informal conversation probably helped inter-resident relationships." – Surgical instrumentation (7.1)* *"Residents tended to sit in groups with [their] specialty, getting to know one another." – Pre-operative preparation (7.2)* *"The rapport between residents and mock patients seemed to be good. After each scenario, they would break character, laugh together about the 'acting' and then start the debrief." – Breaking bad news (7.3)*
Suggestions for future iterations of the boot camp (8)	*"Perhaps an instructor could be appointed to a group or table? Two instructors tended to stay at one table, and the third flowed between the last two, but perhaps the extra instructor could help increase the informal knowledge sharing/question answering for the residents…" – Surgical instruments (8.1)* *"Asking around, people mentioned that they found the scenarios well-made and very realistic. Facilitator XX mentioned that having another scenario with someone under the influence of drugs or alcohol at the time of consent might be helpful as it could help clarify some of the issues around determining if someone is actually capable of providing consent." – Informed consent (8.2)*

Table 5

Process Evaluation Phase: Facilitator Exit Interviews

Theme	**Sub-theme**	**Sample Quotations**
Gap between medical school and residency (9)	Due to varied experiences in medical school (9.1)	*"Medical school education is highly variable, so really, it depends on the individual and their experiences." - Facilitator 5 (9.1.1)* *"I think there is in some students, and I think there is not in others. So I think it's really going to depend on what their comfort is, and the rotations that they've been on, and how much they have been exposed to it." –Facilitator 9 (9.1.2)*
	It is an experience gap, not a knowledge gap (9.2)	*"Like I said I think it's not a knowledge thing [between medical schools], they pass all the same exams, they do the same things, we have the same responsibilities and stuff. I just really think it's an experience." – Facilitator 9 (9.2.1)* *"Medical school teaches you enough to get you into trouble, residency teaches you enough to get you out of trouble, and it takes a few years in practice before you figure out how to not get into trouble in the first place…Yes there's a gap, I am sure there's a gap. But it's not necessarily a knowledge-based gap, it's an experience gap, and it takes time to grow that experience."–Facilitator 4 (9.2.2)*
Surgical Foundations Boot Camp Program (10)	Willingness to return (10.1)	*"Sure! I would do it again."–Facilitator 9 (10.1.1)* *"Oh ya! I like to preach about sepsis [laughter]."–Facilitator 4 (10.1.2)* *"Absolutely [would be willing to come back]."–Facilitators 1,2 & 3 (10.1.3)*

Table 6

Resident Perspectives on the 2016 Surgical Foundations Post-Boot Camp Survey

Theme	Sub-theme	Sample Quotations
Usefulness of the boot camp (11)	Increased feelings of preparedness (11.1)	*"Yes very useful, took my preparedness from a 1 or 2 to a 6." - Resident 1 (11.1.1)* *"Absolutely [useful]. Majority of topics were very high yield and I felt more prepared during my weekend call during boot camp to handle many of the issues we had sessions on." - Resident 8 (11.1.2)*
	Advanced learning (11.2)	*"Very useful to prepare us. Things that we might need months to pick up, the camp helped us know it in advance." – Resident 31 (11.2.1)*
	Leveled the playing field (11.3)	*"Very useful, feels like everyone is at the same level now regardless what school they graduated from." – Resident 18 (11.3.1)*
	Opportunity to meet residents and faculty in a stress-free setting (11.4)	*"YES [useful]! Good way to ease into residency, get to know fellow residents." – Resident 19 (11.4.1)* *"Good to practical technical skills and get to know the residents." – Resident 24 (11.4.2)*
	Clarified expectations (11.5)	*"Good opportunity to cover the expectations of residents." – Resident 26 (11.5.1)* *"The expectations of what standards need to be met as an R1 were clarified." – Resident 17 (11.5.2)*
Aspects of the boot camp that worked well (12)	Boot camp format (12.1)	*"Great format to start with didactic sessions, simulations in small group and regroup for discussion." –Resident 2 (12.1.1)* *"Small groups - put didactic lectures into practice, helps consolidate learning." – Resident 7 (12.1.2)*
	Practical 'hands on' sessions (12.2)	*"All sessions in the surgical skills lab were great, particularly because they were hands-on." – Resident 31 (12.2.1)* *"Suturing & Hand tying -great to have feedback. Electrocautery - good to practice in a low risk environment." – Resident 26 (12.2.2)*
	Mock patient sessions (12.3)	*"[Mock patients] for BBN and Informed Consent sessions felt excellent. They really enhanced the simulations and made the scenarios more realistic and subsequently challenging." - Resident 2 (12.3.1)* *"Mock patients were definitely helpful. However training them could have made it more useful/less easy which would've been better." –Resident 10 (12.3.2)*
	Resident facilitators (12.4)	*"The best sessions were those in which we were able to ask questions to senior residents." – Resident 12 (12.4.1)* *"Resident Q&A sessions were great because they helped reduce anxiety." – Resident 22 (12.4.2)*

Table 6 Cont'd

Theme	Sub-theme	Sample Quotations
Aspects of the boot camp that could be improved (13)	Session timing (13.1)	*"Timing for some sessions was too long, others too short." – Resident 16 (13.1.1)*
	Boot camp was too general surgery focused (13.2)	*"Some sessions were very geared towards general surgery and would have been helpful if the focus had been broadened to consider other specialties." – Resident 16 (13.2.1)*
	Timing of the Advanced Trauma and Life Support (ATLS) course (13.3)	*"In ATLS we do the airway/chest tube/line skills. If ATLS could be incorporated into boot camp during the week this would be helpful." – Resident 6 (13.3.1)*
	Operating room (OR) conduct, tubes/drains, and electrosurgery sessions (13.4)	*"OR conduct - redundant with medical school training and too long, and not clinically relevant." – Resident 26 (13.4.1)* *"Too much focus on airway/chest tubes/central lines for practicing that may never be done for some specialties." – Resident 11 (13.4.2)* *"Electrosurgery lecture: later in the year, hands-on session was too long but helpful." – Resident 28 (13.4.3)*
	Adjust pre-operative preparation scope (13.5)	*"Pre Op Sessions: Anesthesia session felt way too broad and non-specific in scope." – Resident 1 (13.5.1)* *"Pre Op anesthesia was too much info." – Resident 16 (13.5.2)*
Remaining concerns entering residency (14)	Time management (14.1)	*"Managing time and prioritizing tasks while on call, how to stay organized with patient information." - Resident 1 (14.1.1)* *"Managing time during clinic/day pager…" – Resident 20 (14.1.2)*
	Caring for sick patients independently/acute care (14.2)	*"Still nervous about managing really sick patients on my own." – Resident 12 (14.2.1)* *"Remaining calm and collected (focused) when faced with an overwhelming emergency i.e. crashing patient, and no one else around. Running my first code alone…" – Resident 9 (14.2.2)*
	Navigating hospital systems (14.3)	*"Learning how the hospital system/EMR works in Hamilton centers." – Resident 17 (14.3.1)* *"Different computer programs at the different sites." - Resident 22 (14.3.2)*
	Specialty specific knowledge/skills (14.4)	*"Still concerned about specialty skills/knowledge." – Resident 14 (14.4.1)* *"Specialty specific concerns (minor), otherwise I feel pretty prepared." – Resident 25 (14.4.2)*

Table 7

Factors for Successful Curricular Change by Bland et al. (2000) and How They Relate to the McMaster Surgical Foundations (SF) Boot Camp (BC)

Stage of Innovation	Factors for Success	Description	Present in BC	Description
Context	Organization mission and goals	Innovation is compatible with institution's mission, goals, and educational philosophy.	Yes	McMaster University is a student-centered and research-intensive university. McMaster's mission statement describes the institution as being dedicated to creativity, innovation and excellence.
	History of change in organization	Institution has a history of successfully implementing innovations.	Yes	McMaster University has been the leader in previous medical education initiatives such as Problem-Based Learning (PBL).
	Politics: (Internal networking)	Having a strong, influential advocate at the forefront of the change effort. Gaining "buy-in" from powerful individuals from organization.	Yes	The SF Program Director advocated for the BC. Department of Surgery leadership also provided their support early in the project.
	Politics: (Resource allocation)	Ensuring appropriate funding is in place (internal/external).	Yes	The BC was supported by internal funds from the SF program.
	Politics: (External support)	Proposed innovations are linked with interests and needs of external bodies.	Yes	The SF BC outcomes were linked to the RCPSC CBD initiative.
	Politics: (Organizational structure)	Collaboration between departments.	Yes	The SF committee involved Program Directors and Program Coordinators from nine different surgical specialties at McMaster.
Curriculum	Need for change	Widespread agreement that the innovation is relevant to real problems the potential users experience.	Yes	Both residents and faculty at McMaster agreed that learners could be better supported during the transition to residency.

Table 7 Cont'd				
Stage of Innovation	**Factors for Success**	**Description**	**Present in BC**	**Description**
Curriculum	Scope and complexity of innovation	Innovation is large enough to justify the human and financial costs, but feasible to maintain long-term.	Yes	Previous BCs estimated cost at $1000/trainee. The SF budget, including BC, for one year was estimated at $39,000 (not including salaries).
Process	Cooperative climate	Innovation team demonstrates collaborative problem solving, interpersonal respect, support, and cohesion.	Yes	Throughout the development and implementation process, the SF Program Director, SF Program Coordinator, graduate student researcher, and graduate student supervisor/educational scientist worked closely to create a cooperative climate.
	Participation by organization's members	Organization members' participation increased collective ownership and commitment to making it successful.	Yes	Many members of the SF committee provided input on the boot camp schedule and also volunteered as a facilitator.
	Communication	Frequent, timely, and forthright communication/updates among participants and organization members create a unified purpose. Also important to allow dissenters to air objections and invite participation and discussions to collaboratively problem solve.	Yes	Updates on the BC were provided at the monthly SF committee meetings. A variety of methods were used to collect feedback on the innovation (observational notes, post-BC survey, and facilitator interviews). Major themes were summarized and disseminated back to the SF committee and Department of Surgery leadership.
	Human Resource Development: (Training support)	Training support should be ongoing and of high quality to continue cultivation of new leaders. New members should be 'brought up to speed'.	Yes	This was addressed in future iterations of the BC (see epilogue).

Table 7 Cont'd

Stage of Innovation	Factors for Success	Description	Present in BC	Description
	Human Resource Development: (Reward structure)	Incentives to reward participation in the innovation	No	BC facilitators were volunteers. Faculty participation was noted on STAR-CV, an online system for reporting faculty activities at McMaster. However, no other rewards were provided.
	Human Resource Development: (Evaluation)	Legitimizes the innovation process by holding it to standards of analysis that faculty regard as valid and meaningful	Yes	The graduate student researcher followed the CIPP program evaluation model throughout the innovation.
	Performance dip	To manage decline in organizational performance, acknowledging losses associated with the old curriculum, celebrate successes, make minor adjustments and continue monitoring progression towards the goal	Yes	This was addressed in future iterations of the BC (see epilogue).
	Leadership	Leaders must be assertive, flexible, able to mobilize others, and effectively communicate their vision of curricular change	Yes	SF Program Director and Department of Surgery Associate Chair of Education were advocates for the BC program.

Table 8

Findings From the 2016 Boot Camp and Subsequent Changes to the 2017 Program

Theme	Description	Major Changes
Session timing	Residents suggested many sessions were too long and could be condensed to make room for other content. Observers also noted that sessions felt long and resident attention had often decreased by the end of the morning sessions.	Length of electrosurgery, operating room conduct, and handover was reduced. Session such as navigating electronic platforms, moving to Hamilton and resident wellness, and residents as teachers, and draping were added in the extra time. Boot camp start time was changed from 0800 to 0900am each day.
Session overlap	Residents felt there was significant overlap with the tubes and drains session and the ATLS course. Residents also felt the ATLS course should occur during the workweek rather than during their personal time on the weekend.	ATLS was moved from the weekend to the last 2 days of the BC. The 2 tubes and drains sessions, 1 of 4 suturing sessions, and the open-practice day were removed to make space for ATLS during the week.
Desire for specialty-specific information	Residents wished for specialty-specific content in the BC. At the end of the BC, residents said this was one of the areas where they were still concerned.	A 2-hour block for specialty-specific activitie was added to the BC. Program Directors wer made aware and it was up to the individual specialties to schedule programming during this time.
Fidelity of mock patient activities	Residents suggested providing more variability in the cases (less general surgery focus) and training for the mock patients would improve the small group breakout activities during the breaking bad news and informed consent sessions	Asked Program Directors from all the 9 surgical specialties involved in SF to provide scenarios Volunteer mock patients were required to complete 1 hour of training before the sessions Mock patient sessions were moved to the center for simulation-based learning to increase fidelity
Value of resident-led sessions	Residents and observers suggested the resident-led sessions were extremely valuable. However resident-facilitators said it was difficult to get away from their clinical duties to participate in BC	Resident-led sessions were moved to Wednesday mornings to promote resident-facilitators. Wednesday mornings are the academic half-day for most programs, thus resident-facilitators would not be missing clinical duties
Information overload	Residents suggested there was too much information presented in the didactics, particularly the pre-operative preparation and electrosurgery sessions. Facilitators echoed these concerns	Session objectives were redesigned. A case-based discussion was added to pre-operative preparation so residents could consolidate their knowledge. Lab component of electrosurgery session was redesigned to include multiple different stations.

Table 9

Findings From the 2017 Boot Camp and Subsequent Changes to the 2018 Program

Theme	Description	Major Changes
Session timing	Residents wished sessions including information on hospital logistics were scheduled during the first week of BC, as many were on call over the weekend.	Navigating electronic platforms, ward management, and "what I wish I had known" were all moved to the first week of the BC schedule.
Session overlap	Residents suggested the instrumentation, scrub session, sharps and aseptic technique, and sterility sessions had some redundancies and could be shortened/ combined. Residents also suggested the residents as teachers session had some overlap with the online modules from postgraduate medicine.	One afternoon session was created where residents rotated through a scrub session station, gowning and gloving station, and introduction to instruments station. This addressed the overlap and also meant nurse educators only had to attend the BC for one afternoon. Resident as teachers session objectives were redesigned to focus on practical tips for supervising medical students in the clinical environment.
Desire for specialty-specific information	Few specialties used the 2-hour block for specialty-specific time. Thus, residents still wanted more specialty-specific content during the BC.	The draping session required at least one facilitator from each surgical specialty. Thus, specialty-specific time was scheduled immediately following the draping session in the hopes that the facilitators would be motivated to stay and utilize that time.
Fidelity of mock patient activities	Residents felt the mock patient activities were extremely useful, however often they did not have the clinical knowledge to perform the task (taking consent or breaking bad news) appropriately. Having a short period of time to read the scenario felt unrealistic. Observers also noted that a number of volunteer mock patients cancelled the day of, thus the sessions were a bit disorganized and did not go according to plan.	Residents were encouraged to review the scenarios the night before to ensure they were comfortable with the clinical aspect of the case. Standardized patients were hired in place of volunteer mock patients.
Resource availability	Residents wished that the BC PowerPoints were provided ahead of the sessions so they could prepare. Residents also wished the resources on hospital logistics were summarized into a handbook they could carry.	Facilitators were instructed to provide their PowerPoints ahead of the BC. One of the resident-facilitators offered to create a handbook with major resources.

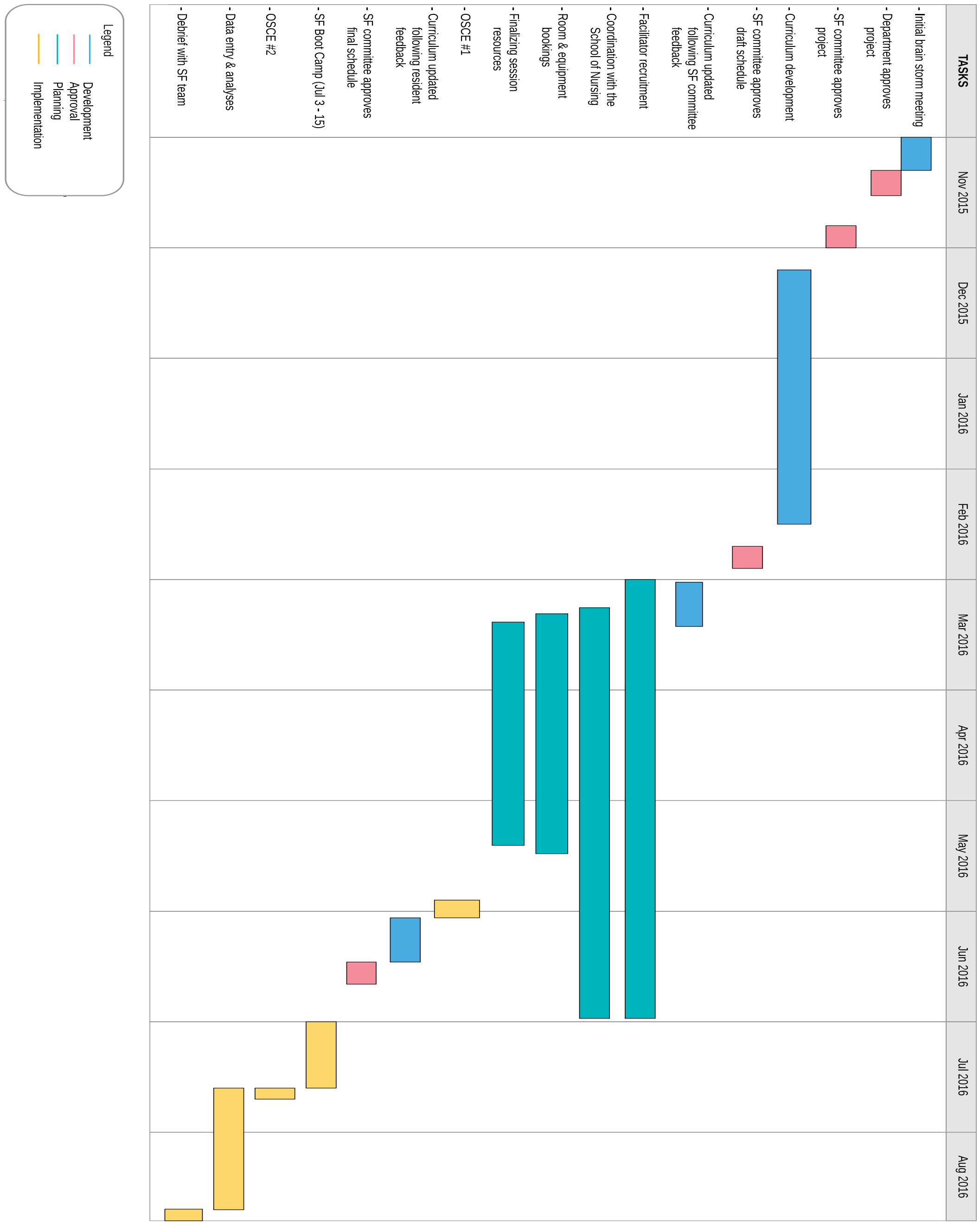


Figure 1. Boot Camp Development and Implementation Timeline

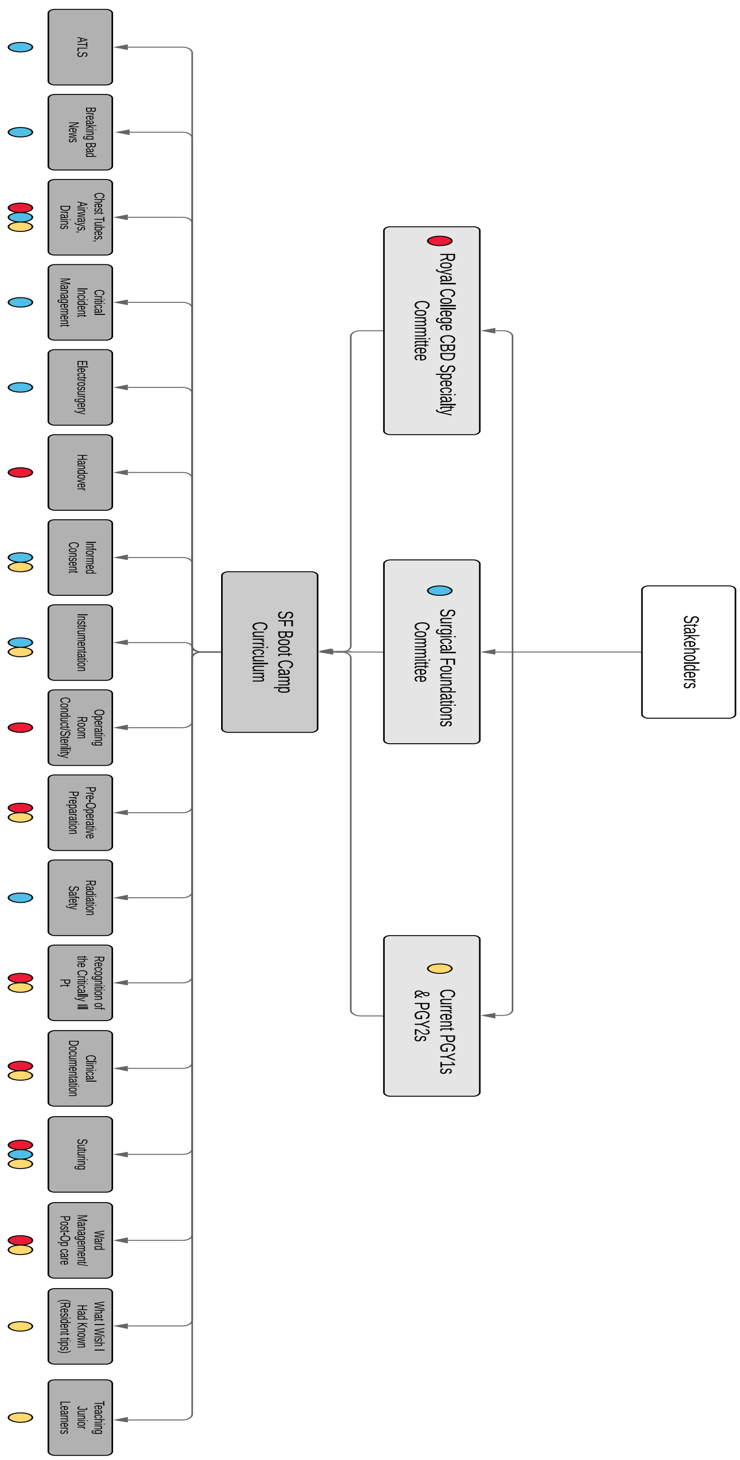


Figure 2. Boot Camp Content Development Map

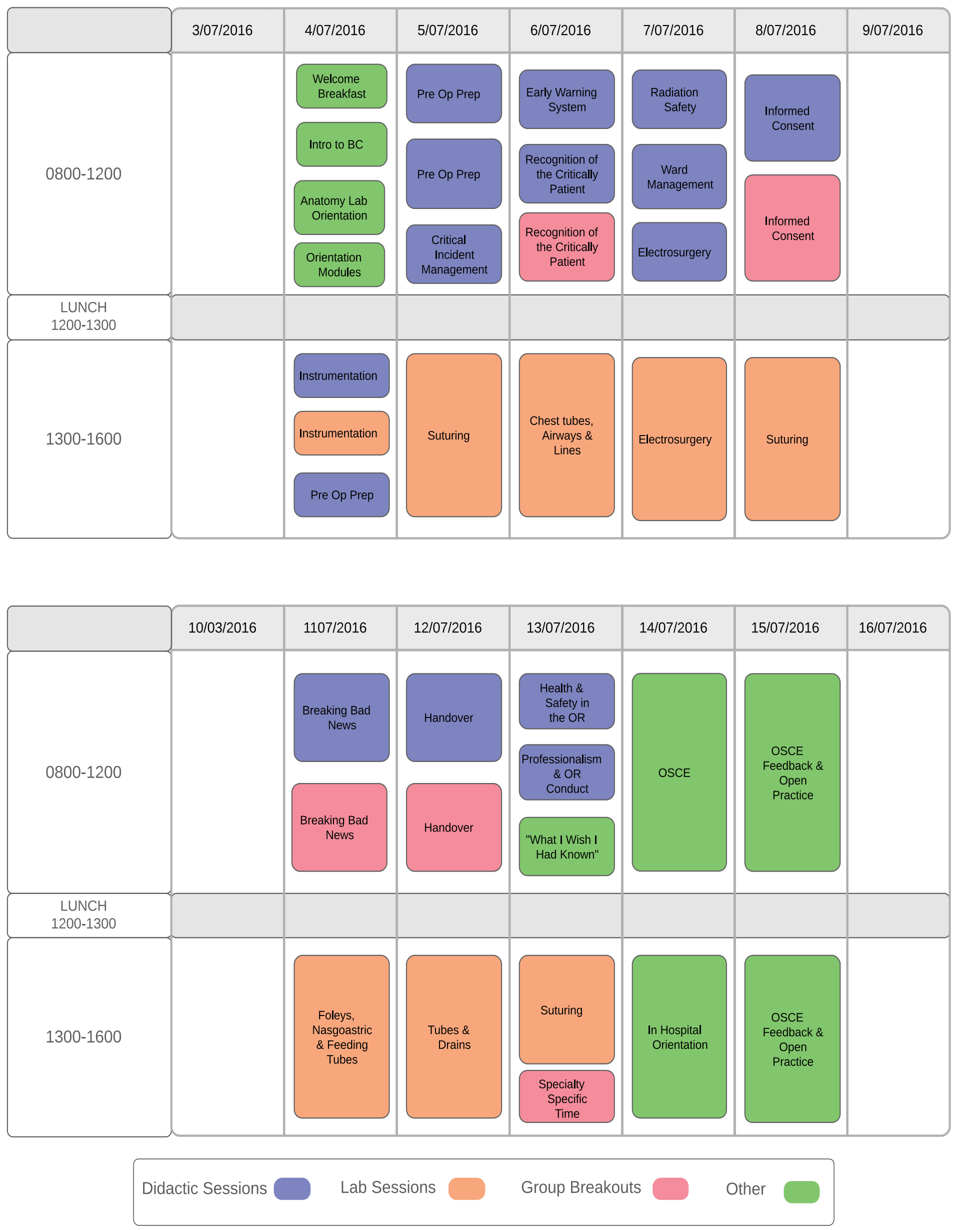


Figure 3. 2016 McMaster Surgical Foundations Boot Camp Schedule

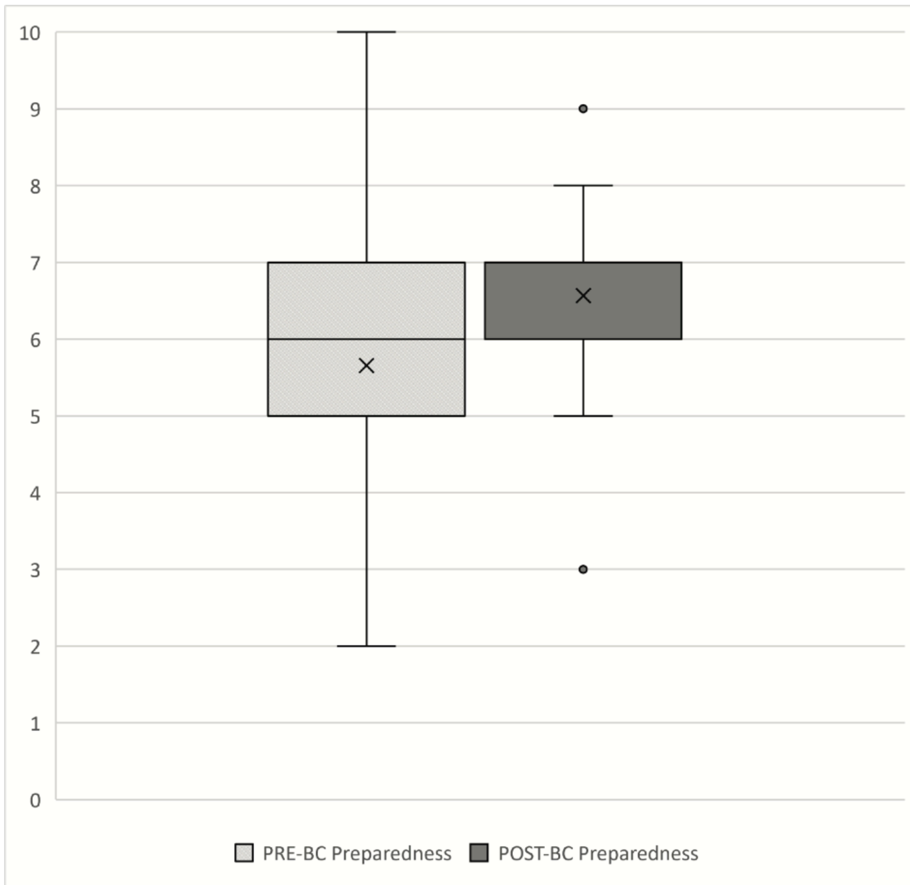


Figure 4. Resident Self-Rated Preparedness Scores Pre- and Post-Boot Camp

Appendix A

Survey for Current Surgical Foundations Trainees (June 2016)

Part I: Surgical Foundations Curriculum

1. How well did medical school prepare you for residency? Please circle below [1 = did not prepare at all; 10 = completely prepared]

1 2 3 4 5 6 7 8 9 10

2. Were there any skills you were expected to know how to do, but did not feel prepared/confident performing in your first two weeks of training? Please list/explain.

3. Are there any skills you currently are expected to know how to do, but do not feel prepared/confident to perform? Please list/explain.

4. What skills would be useful to include in the Surgical Foundations boot camp (occurring the first 2 weeks of residency for incoming PGY1s)?

5. Do you have any other suggestions for the Surgical Foundations boot camp?

6. Do you have any suggestions for improving the Surgical Foundations year-round curricula?

7. Please rate the overall value of Surgical Foundations [1 = not useful at all; 10 = extremely useful]

1 2 3 4 5 6 7 8 9 10

8. Please list any other comments.

TURN TO NEXT PAGE

Part II: Assessment in Surgical Education

1. How often did you personally complete assessments/evaluations of someone (or something) else in Surgical Foundations?

2. How often did you personally complete assessments/evaluations of someone (or something) else in your home program?

3. How often are you assessed/evaluated in Surgical Foundations?

4. How often are you assessed/evaluated in your home program?

5. Do these evaluations provide you with meaningful feedback to help you progress in the program? Please explain.

6. Do you have any suggestions for improving assessment/evaluation in Surgical Foundations?

END OF SURVEY

Appendix B

Focus Group Script for Current Surgical Foundations Trainees (June 1st 2016)

Part I: Surgical Foundations Curriculum

1. How well did medical school prepare you for residency?

2. What skills would be useful to include in the new Surgical Foundations boot camp (occurring the first 2 weeks of residency for incoming PGY1s)?

3. Do you have any other suggestions for the Surgical Foundations boot camp?

4. Do you have any suggestions for improving the Surgical Foundations year-round curricula?

5. What would you rate the overall value of Surgical Foundations [1 = not useful at all; 10 = extremely useful]

6. Do you have any other comments about Surgical Foundations?

Part II: Assessment in Surgical Education

1. How often are you assessed/evaluated in Surgical Foundations?

2. [Follow Up] What about your home program?

3. Do these evaluations provide you with meaningful feedback to help you progress?

4. Do you have any suggestions for improving assessment/evaluation in Surgical Foundations?

5. [Follow Up] What about your home program?

6. Are you aware of the new competency-based curriculum rolling out in the next few years?

7. [Follow Up] If so, what are your thoughts on the new curriculum?

END OF SCRIPT

Appendix C

Prompts for daily observation (SF BC July 2016)

1. Did the session start on time?

2. Did the correct (assigned) facilitator show up?

3. Was it a "good" session? Why or why not?

4. In general what happened during the session?

5. Did the residents seem engaged? Were they asking questions?

6. What types of questions were residents asking (task specific, broader context, specialty specific)?

7. How were the residents interacting with each other?

8. What worked well in the session? Why?

9. What didn't work well in the session? Why?

10. Any additional comments?

Prompts adapted from: Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories AMEE Guide No. 67. Medical Teacher, 34, e288-e299.

Appendix D

Facilitator Exit Interview Script (July 2016)

Hello, I am [Interviewer's name], I am a [interviewer's role] with the Sonnadara Lab. I was wondering if you would be willing to participate in a brief interview regarding the boot camp session you facilitated and the transition between medical school and residency more broadly? The results will be used to inform future iterations of the boot camp.

Would you be willing to be recorded? All names will be removed and just general themes will be shared.

1. Do you think there is a gap between medical school & residency? If so, why?
2. What would you change about this session next year?
3. Would you be interested in coming back?

Thank you for participating, we really appreciate it.

END OF SCRIPT

Appendix E

Pre –Boot Camp Resident Survey (July 2016)

Part I: Surgical Foundations Curriculum

1. How prepared do you feel for residency? Please circle below [1 = not prepared at all; 10 = completely prepared]

1 2 3 4 5 6 7 8 9 10

2. What (if any) concerns do you have about entering residency? Please list/explain.

3. Do you think including a boot camp in the first two weeks will be useful for incoming residents? Please explain.

4. What skills/information would be useful to include in the Surgical Foundations boot camp?

5. Please list any other comments regarding the SF boot camp.

Part II: Assessment in Surgical Education

1. How often did you personally complete assessments/evaluations of someone (or something) else during clerkship?

2. How often were you assessed/evaluated during clerkship?

3. Did those evaluations provide you with meaningful feedback to help you progress towards residency? Please explain.

4. Do you believe regular assessments should be part of residency training? Please explain.

Part III: Competence By Design

1. Were you aware the Royal College is switching to a competence by design (CBD) residency curriculum in the coming years?

2. If yes, what do you know about CBD? Please explain.

Appendix F

Post- Boot Camp Resident Survey (July 2016)

Part I: Surgical Foundations Curriculum

1. How prepared do you feel for residency? Please circle below [1 = not prepared at all; 10 = completely prepared]

   1 2 3 4 5 6 7 8 9 10

2. What (if any) concerns do you still have about entering residency? Please list/explain.

3. Was including a boot camp in the first two weeks useful for incoming residents? Please explain.

4. What sessions (if any) worked well? Please explain why.

5. What sessions (if any) could be improved? Please explain why/how.

6. What (if any) content did you feel was missing?

7. What (if any) content could be removed?

8. Did you find the source documents useful? Please explain why or why not.

9. Did you find working volunteers as "SPs" useful? Why or why not.

10. Please list any other comments regarding the SF boot camp.

TURN TO NEXT PAGE

Part II: Assessment in Surgical Education

1. How often did you personally complete assessments/evaluations of someone (or something) else during boot camp?

2. How often were you assessed/evaluated during boot camp?

3. Did those evaluations provide you with meaningful feedback to help you progress towards residency? Please explain.

4. Do you believe regular assessments should be part of residency training? Please explain.

END OF SURVEY

Appendix G

Sample OSCE Assessment Form: Informed Consent

Instructions to Candidates:
A 65 y/o male 5 days post-op from an open sigmoid resection with primary anastomosis for diverticular disease has developed a wound infection with a 5 x 3 cm area of erythema and fluctuance. There is no discharge from the wound and the skin staples are still in place. Please obtain informed consent from the patient to incise and drain the wound abscess.

ITEM	**Not done, Incorrect**	**Done, Correct**
1. Explained the rationale for the procedure	**0**	**1**
2. Accurately described the procedure	**0**	**1**
3. Described the risks of the procedure	**0**	**1**
4. Described the benefits of the procedure	**0**	**1**
5. Confirmed the patient understood the procedure	**0**	**1**
6. Asked if the patient had any questions	**0**	**1**
7. Correctly completed both consent forms	**0**	**1**

MAXIMUM TOTAL SCORE **/7**

Please circle the number corresponding to the candidate's performance regardless of the candidate's level of training.

Knowledge of Procedure				
1	**2**	**3**	**4**	**5**
Deficient knowledge; required assistance to explain steps of procedure		Knew most steps of the procedure		Explained all steps of the procedure without prompting
Communication				
1	**2**	**3**	**4**	**5**
Difficulties verbalizing information; lacked clarity		Adequately verbalized information		Clearly and accurately verbalized information
Professionalism				
1	**2**	**3**	**4**	**5**
Did not listen to the patient's concerns		Listened to most of the patient's concerns		Listened to all of the patient's concerns
Entrustment				
1	**2**	**3**	**4**	**5**
Requires complete supervision/guidance		Requires some supervision/prompting		Requires no supervision, ready for independent practice
Overall Performance				
1	**2**	**3**	**4**	**5**
Demonstrated no competence; requires extra support		Demonstrated some competence; remains in training		Demonstrated complete competence; ready for independent practice

3.8 References

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-367.

Blackmore, C., Austin, J., Lopushinsky, S. R., & Donnon, T. (2014). Effects of Postgraduate Medical Education "Boot Camps" on Clinical Skills, Knowledge, and Confidence: A Meta-Analysis. *J Grad Med Educ, 6*(4), 643-652. doi:10.4300/JGME-D-13-00373.1

Bland, C., Starnaman, S., Wersal, L., Moorhead-Rosenberg, L., Zonia, S., & Henry, R. (2000). Curricular Change in Medical Schools: How to Succeed. *Acad Med, 75*, 575 - 594. Retrieved from http://ovidsp.dc1.ovid.com.libaccess.lib.mcmaster.ca/.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Creswell, J. W. (2007). *Qualitative Research Inquiry Methods; Second Edition*. Thousand Oaks, CA: SAGE Publications.

Cypress, B. S. (2017). Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations. *Dimens Crit Care Nurs, 36*(4), 253-263. doi:10.1097/DCC.0000000000000253

Flexner, A. (1910). *Medical Education in the United States and Canada: A report to the Carnegie Foundation for the Advancement of Teaching*. New York City, NY: The Carnegie Foundation for the Advancement of Teaching.

Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories AMEE Guide No. 67. *Medical Teacher, 34*, e288-e299. doi:10.3109/0142159X.2012.668637

Gandomkar, R. (2018). Comparing Kirkpatrick's original and new model with CIPP evaluation model. *Journal of Advances in Medical Education & Professionalism, 6*, 94-95. Retrieved from https://journals-scholarsportal-info.libaccess.lib.mcmaster.ca/pdf/.

Harden, R. M., Stevenson, M., Wilson Downie, W., & Wilson, G. M. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal, 1*, 447-451. Retrieved from https://www-jstor-org/.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qual Health Res, 15*(9), 1277-1288. doi:10.1177/1049732305276687

Khan, M. E., Anker, M., Patel, B. C., Barge, S., Sadhwani, H., & Kohle, R. (1991). The use of focus groups in social and behavioural research: some methodological issues. *World Health Stat Q, 144*(3), 145-149. Retrieved from https://www.researchgate.net/.

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*, 273-278. doi:10.1002/bjs.1800840237

McMaster University, D. o. S. (2017). Surgical Foundations. *Education.* Retrieved from https://surgery.mcmaster.ca/education/surgical-foundations

Morrison, E. W. (1993). Longitudinal Study of the Effects of Information Seeking on Newcomer Socialization. *Journal of Applied Psychology, 2*, 173-183. Retrieved from https://www.researchgate.net/.

Neergaard, M. A., Olesen, F., Andersen, R. S., & Sondergaard, J. (2009). Qualitative description - the poor cousin of health research? *BMC Med Res Methodol, 9*, 52. doi:10.1186/1471-2288-9-52

NVivo. (2019). What is NVivo? Retrieved from https://www.qsrinternational.com/nvivo/what-is-nvivo

Sandelowski, M. (2000). Whatever happened to Qualitative Description? *Research in Nursing & Health, 23*, 334-340. Retrieved from https://journals-scholarsportal-info.libaccess.lib.mcmaster.ca/.

Sandelowski, M., & Leeman, J. (2012). Writing Usable Qualitative Health Research Findings. *Qual Health Res, 22*, 1404-1413. doi:10.1177/1049732312450368

Stufflebeam, D. L. (1971). The Relevance of the CIPP Evaluation Model for Educational Accountability. *Paper at the Annual meeting of the American Association of School Administration*, 30. Retrieved from https://files.eric.ed.gov/fulltext/ED062385.pdf.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models and applications*: Jossey-Bass.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nurs Health Sci, 15*(3), 398-405. doi:10.1111/nhs.12048 van der Vleuten, C., Dolmans, D. H. J. M., & Scherpbier, A. J. J. A. (2009). The need for evidence in education. *Medical Teacher, 22*(3), 246-250. doi:10.1080/01421590050006205

Van Maanen, J., & Schein, E. H. (1979). Toward A Theory of Organizational Socialization. *Research in organizational behaviour, 1*, 209-264. Retrieved from https://dspace.mit.edu/bitstream/handle/1721.1/1934/?sequence=1.

Wagoner, H., & Seltz, B. (2019). Attending Physicians' Perspectives of Resident Academic Half Day. *Teach Learn Med*, 1-9. doi:10.1080/10401334.2018.1551140

Young, J. Q., Ranji, S. R., Wachter, R. M., Lee, C. M., Niehaus, B., & Auerbach, A. D. (2011). "July Effect": Impact of Academic Year-End Changeover on Patient Outcomes. *Annals of Internal Medicine, 155*, 309-315. Retrieved from https://www.annals.org/.

Chapter 4

Long-term Effectiveness of a Boot Camp for Novice Surgical Trainees:

A Convergent Mixed Methods Study

4.0 Long-term Effectiveness of a Boot Camp for Novice Surgical Trainees: A Convergent Mixed Methods Study

4.1 Introduction

As discussed in chapter three, evaluating and understanding the long-term implications of an educational initiative is imperative for successful curricular change. Long-term outcome data can help stakeholders identify intended and unintended consequences of a program, inform changes, maintain momentum, and secure resources for sustainability (Bland et al., 2000). To date, there remains a paucity of literature on the long-term implications of boot camps (BCs) in medical training.

In a systematic review of studies exploring the effectiveness of BC programs in medical training, Blackmore, Austin, Lopushinsky, and Donnon (2014) identified 15 articles that met their inclusion criteria. Of these studies, 13/15 (87%) used repeated measures, pretest-posttest designs to capture changes in residents' knowledge, self-confidence or performance of clinical skills over the course of a BC. Six of the 15 articles (40%) also compared knowledge and/or performance of clinical skills between learners who had completed a BC with a control group who had not completed the BC. While the results suggest that BCs can improve residents' skills, knowledge and confidence, all studies included in the review presented data collected immediately following their BC program.

To the best of our knowledge, there has only been one study to date exploring longer-term effectiveness of a BC program. A study by Sonnadara et al. (2012) evaluated residents' performance on an Objective Structured Clinical Examination (OSCE) immediately following a BC, and again at seven months into training. Results from the Sonnadara et al. (2012) study suggest that their BC cohort scored significantly higher than the traditionally trained cohort immediately following the BC. Moreover, the BC cohort maintained their performance advantage at seven months in training. From this set of results, the authors conclude that BCs can improve basic skills, and that those skills have an excellent retention rate (Sonnadara et al., 2012). Sonnadara et al. (2012) go on to suggest that BCs have the potential to accelerate later clinical learning as less time and attention need to be dedicated to the basic skills. While these results are promising, we still have no information on the impact of BCs beyond seven months into training.

As improving patient care is the main motivator for initiatives in medical education (van der Leeuw, Lombarts, Onyebuchi, & Heinerman, 2012), ideally we could use patient-related outcomes to measure the long-term impact of BC programs. However, there are a number of reasons why patient-related outcomes are not considered a good measure of resident performance (Brydges, Hatala, Zendejas, Erwin, & Cook, 2015). Many healthcare systems are set up so residents, attending physicians, pharmacists, nurses, and many other allied health professionals, all work together as a team to ensure high-quality patient care. With this structure in mind, patient outcomes are often more reflective of team performance than the performance of an individual resident. Secondly, there are a number of patient-specific factors, such as comorbidities or adherence to treatment, that are not in the control of the resident or healthcare team. It can be difficult to separate the influence of those factors from the performance of the resident on patient-related outcomes. Lastly, what trainees encounter in the clinical environment is highly unpredictable (recall chapter 3.4.1). If the intention is to compare performance between residents who received an educational intervention (such as a BC), and those who did not, there is no guarantee that all residents will have had similar previous clinical encounters.

In light of the challenges with using patient outcomes, most studies to date, including the study by Sonnadara et al. (2012), have used quantitative measures of performance, such as Objective Structured Clinical Examinations (OSCEs), to capture performance change in residents who completed a BC. The OSCE was developed in 1975 to improve the measurement of competence in medical trainees (Harden, Stevenson, Wilson Downie, & Wilson, 1975). The original OSCE consisted of eight pairs of couplet stations, five-minutes each. In the first station, medical trainees were given written instructions and asked to complete a task or procedure - for example, take a history or physical examination from a mock patient. In the second station, medical trainees had to answer questions on their findings and interpretations from the previous station. Trainees were scored on a checklist for each station. The final OSCE score was a combination of checklist scores from all stations. In contrast to patient-related outcomes, this OSCE format standardized assessment of medical trainee performance, and allowed multiple learners to be evaluated on the exact same scenario. The OSCE provided educators with a way to isolate trainee performance from all other variables and compare performance between residents (or intervention cohorts) with no risks to patient safety. As a result, the OSCE gained considered traction amongst medical educators and is now considered the gold standard for assessment of resident performance in medical training (Pugh, Touchie, Humphrey-Murto, & Wood, 2016).

While OSCEs seem to be the best quantitative measure for capturing resident performance over time, we know from chapter three that there are important factors relating to BCs that performance tests, such as the OSCE, fail to capture. For example, qualitative data in chapter three suggests the utility of BCs is impacted by variability in medical school experiences; that BCs are not able to fully prepare learners for the transition to residency, rather they can provide residents with basic knowledge, tips, resources, and coping strategies for things they will encounter in the clinical environment; and that there is a large social benefit to BCs that eases the transition and perhaps sets up later clinical success. Capturing this type of qualitative data seems critical to understanding the true impact of BC programs. Yet, there have been very few qualitative studies on BC effectiveness. Of the studies that exist, all have focused on short-term outcomes of BC programs.

The purpose of this study was to collect *both* qualitative and quantitative data over two years to capture a more comprehensive picture of what BCs are effective in improving, and what contextual factors contribute to that effectiveness. To accomplish this goal, we used a Convergent Parallel Mixed Methods (CPMM) design.

Mixed methods research involves purposefully collecting, analyzing, and integrating both quantitative and qualitative data in response to a research question (Creswell & Plano Clark, 2018). Typically, quantitative methods are objective and generalizable, but lack information on the context. Qualitative methods highlight the contextual piece, but are criticized for their subjectivity and lack of generalizability. Mixed methods are meant to emphasize the strengths of both quantitative and qualitative methods, while offsetting their weaknesses (Creswell & Plano Clark, 2018). In the CPMM design, qualitative and quantitative data are collected in parallel, analyzed separately, and then merged (Creswell & Plano Clark, 2018). The CPMM design was one of the first mixed methods designs described, and has many attractive qualities (Creswell & Plano Clark, 2018). Firstly, the quantitative and qualitative strands are analyzed independently prior to integration, emphasizing robust methods practice and giving equal weight to both sources of data. Convergent parallel mixed methods (CPMM) designs facilitate the direct comparison of rich, qualitative data with statistical inferences. Convergent parallel mixed methods (CPMM) designs can also be highly efficient as both quantitative and qualitative data can be collected simultaneously.

Using a CPMM design enabled us to explore whether a BC at the onset of residency was effective in improving long-term resident performance, what aspects of performance improved, and what contextual factors contributed to that effectiveness in an efficient and robust manner. The quantitative strand compared OSCE performance between those who completed the BC, and the previous cohort of residents who completed traditional training (without a BC), at one year and two years into training. The qualitative strand explored resident and Program Director perceptions of the usefulness of BCs immediately following the BC and six months into training. Findings from both strands were then integrated through a joint display table to create a comprehensive picture of long-term BC effectiveness.

4.2 Methods

4.2.1 Quantitative Strand

The quantitative strand of this study used a quasi-experimental design to evaluate whether there were significant differences in OSCE performance between residents who completed a BC at the onset of residency, and residents who completed traditional training (TT), at one year and two years into training.

Setting and participants.

The quantitative component of this study took place at McMaster University in Hamilton, Ontario. Participants were recruited from the 2015 cohort of Surgical Foundations (SF) residents who completed traditional training (TT cohort; n = 31), and the 2016 cohort of SF residents who participated in the BC described in chapter three (BC cohort; n = 32). Throughout the academic year (July 2015 – June 2016), the TT cohort received 100, 1.5 hour teaching sessions through an academic half-day format, 1x/week (150 hours). The remainder of their time (52 weeks, which included up to four weeks of vacation time) was spent on clinical rotations. The BC cohort received a BC the first two weeks of July 2016 (63 hours). Following the BC, they completed 92, 1.5 hour teaching sessions through an academic half-day format, 1x/week (138 hours). The remainder of their time (50 weeks, which included up to four weeks of vacation time) was spent on clinical rotations.

Study design.

Annual OSCEs are a component of the SF training program at McMaster. However, in order to be able to use these data for research purposes, the research team obtained consent from the SF program to access and de-identify the assessment data. The research team also obtained verbal consent to use the scores from residents completing the OSCEs.

The SF OSCEs took place at the end of each academic year (June) for the TT and BC cohorts. The BC cohort also completed one immediately following the BC (Table 1). Each OSCE consisted of nine stations: eight skill-based stations and one rest station. The eight skill-based stations drew upon tasks outlined by the Royal College of Physicians and Surgeons of Canada (RCPSC) that residents in SF must be able to do within their first stage of training (SF Transition to Discipline EPAs – see chapter 1.7.5). The eight skill-based stations included: informed consent, pre-operative preparation, suturing, breaking bad news, operating room conduct, recognition of the critically ill patient, handover and hand ties, as well as a multiple-choice quiz on surgical instrumentation, electrosurgery, and radiation safety. While the station topics remained the same for each OSCE cycle, the scenarios were changed to avoid test-retest bias, or an inflation of scores due to the same individuals retaking the same test (Lievens, Buyse, & Sackett, 2005). All OSCE scenarios were created by senior residents or faculty and reviewed by the SF Program Director to ensure equal difficulty between OSCEs. Volunteer examiners (faculty, fellows, and senior residents from the Department of Surgery) evaluated resident performance at each station using task-specific checklists and global-rating scales of entrustment. Our OSCE assessment tools were based on work by Martin et al. (1997), where the authors created dichotomous task-specific checklists for each examination station, and seven 5-point global rating scales common to all stations. The Objective Structured Assessment of

Technical Skills (OSATS) scoring system developed by Martin et al. (1997) is used for many assessments in medical education. In our study, the OSATS-style task-specific checklists were modified to match our OSCE stations, while the global rating scales were modified to reflect the movement in medical education towards using trust to gauge residents' competence (Hauer et al., 2014; Rekman, Gofton, Dudek, Gofton, & Hamstra, 2016) (Appendix A). The Objective Structured Assessment of Technical Skills (OSATS) scoring system is discussed in more detail in chapter five.

Analysis.

Evaluations were completed by the volunteer examiners using pen and paper during the OSCEs and then entered into a secure database by the research team. Scores from the task-specific checklists were converted to percentages to allow for comparisons across stations. The global ratings of entrustment were also compared across stations.

A one-way repeated measures Analysis of Variance (ANOVA) was conducted to evaluate the BC cohort's OSCE performance across time (post-BC, one year into training, two years into training; Table 1). Separate, two-way repeated measures ANOVAs were used to examine whether there were significant differences in OSCE scores between the TT and BC cohorts at one year and two years into training. For the two-way ANOVAs, group was the between subjects factor at two levels (BC cohort and TT cohort), and station was the within subjects factors (see previous station list). All analyses were conducted for checklist scores (converted to percentages) and global ratings of entrustment. All statistical analyses were completed in IBM SPSS (Version 25).

4.2.2 Qualitative Strand

The qualitative strand used qualitative description (QD) to explore resident and Program Director perceptions of the usefulness of BCs immediately following the BC and six months into training. As discussed in chapter 3.2.3, QD is one of many approaches to qualitative research (Creswell, 2007). Qualitative Description is not driven by pre-existing theories and does not try to generate a theory (Sandelowski, 2000). Rather, QD provides a rich, comprehensive summary of different perspectives on one event or experience based on the data collected (Neergaard, Olesen, Andersen, & Sondergaard, 2009). As the goal of the qualitative strand was to explore resident and Program Director perceptions on the usefulness of the BC and whether resident performance changed over time, QD was selected as the most appropriate approach.

Setting and participants.

As per the quantitative strand, the qualitative strand took place at McMaster University in Hamilton, Ontario. Purposeful samples of participants were recruited, including: 1) the cohort of SF residents who completed the 2016 BC at the onset of training; and 2) Program Directors of the surgical residency programs at McMaster University who were part of the SF program in 2016. These two groups were recruited, as they were likely to possess different perspectives on the usefulness of BCs. Moreover, Program Directors are responsible for resident education and assessment, thus their opinion on resident performance was important. The qualitative component of this study was voluntary and the research team obtained verbal consent from all participants.

Study design.

Resident post-boot camp focus group.

Residents who participated in the 2016 BC (n = 32) were invited to join a focus group over their lunch hour on the last day of the BC. Participants were split into four separate focus groups. Volunteer graduate student researchers facilitated the focus groups, as they did not have any affiliations with the SF program or BC. Focus groups were selected as they can provide in-depth insight into participant perception, attitudes, and beliefs in an efficient manner (Khan et al., 1991). Focus groups allow members to interact while sharing their experiences, and allow the researcher to explore themes that go beyond what was included on the focus group script (Khan et al., 1991). Focus groups are also an efficient approach when there is only one opportunity to speak with participants (Cohen & Crabtree, 2006). All focus groups followed a semi-structured script (Appendix B). A semi-structured focus group script was used so the four focus group facilitators would cover the same content, while also pursuing topic trajectories when appropriate. All focus groups were recorded, transcribed verbatim and entered into NVivo (Version 12) for analyses.

Resident six month follow-up survey.

Six months following the BC, the research team approached all residents who participated in the BC (n = 32) during one of their SF academic half-days. Residents were invited to complete a follow-up survey regarding their experiences in the BC, and provide input on the changes we had made to the 2017 BC schedule based on their feedback (Appendix C). The purpose of this survey was to explore whether residents' perceptions of the usefulness of the BC had evolved over time. Surveys were selected as they provide participants with the opportunity to provide their opinions anonymously, promoting candor. Surveys were also a way to ensure we captured as many perspectives as possible without taking too much time away from their academic half-day or clinical responsibilities. Participants were provided with paper copies of the survey and asked to place the completed surveys in a file folder for the research team to collect before they left the academic half-day. All survey responses were anonymous and were coded into Excel sheets and later entered into NVivo for analysis. Participation in both the post-BC focus groups and the six-month follow-up survey was voluntary and did not impact the residents' standing/progression in SF.

Program Director six-month follow-up interviews.

All Program Directors who were part of the SF program in 2016 (n = 9) were contacted by email six months following the BC and asked to participate in a semi-structured interview. A six-month follow-up period was selected, as we were interested in whether or not Program Directors noticed a difference in the clinical performance of residents who had completed the BC compared to previous cohorts who did not complete a BC. We also wanted to explore their perceptions of the usefulness of the BC prior to the next iteration of the BC (July 2017). To address these questions, the interview guide consisted of 12 open-ended questions (Appendix D). One-on-one interviews were conducted by the graduate student researcher (N.W.) or a research assistant (J.W.), and took place at the time and place of the Program Directors' preference. All interviews were recorded and transcribed verbatim. All interview data were de-identified and entered into the NVivo software for analysis.

Analysis.

Qualitative content analysis (QCA) was used to analyze all data collected in the qualitative strand. As discussed in chapter 3.2.4, QCA is generally used with study designs aiming to describe an event or experience from multiple viewpoints. Furthermore, this approach is appropriate when existing theories explaining the event or experience are limited. Specifically, conventional content analysis, a subtype of QCA, was used for this study. Conventional content analysis allows the categories/themes to be based in the data, rather than creating categories based off prior knowledge (Hsieh & Shannon, 2005). Conventional content analysis works best when questions and probes are open-ended. As the main purpose of this study was to describe the perceptions of the usefulness of the BC from multiple perspectives over time, the conventional content analysis method was deemed appropriate for all evaluation phases.

As discussed in chapter 3.2.4, conventional content analysis starts by the researcher reading all of the data repeatedly to gain an understanding of the whole picture (Hsieh & Shannon, 2005). Once that has been achieved, the researcher organizes the data through open coding. Open coding is the process of writing notes or highlighting specific words while reading the data, and then using those notes/words to generate categories. Categories are grouped into themes based on how they are related or linked. Themes are used to create a general description of the event/experience, using language that is relatable and true to the study participants' perspectives. The graduate student researcher (N.W.) completed this process for all qualitative data.

As the post-BC focus groups involved multiple residents in four separate focus groups, a second graduate student researcher (N.C.) coded the focus group transcripts to ensure trustworthiness of the findings. As discussed in chapter 3.2.4, trustworthiness is the accuracy, authenticity, or truthfulness of the data in qualitative research (Cypress, 2017). Trustworthiness is considered equivalent to the empiricist concepts of validity and reliability. Components of trustworthiness include credibility, dependability, and transferability (Cypress, 2017). Comparing themes between two coders is one way to establish credibility, or the accurate representation or truthfulness of data (Cypress, 2017). Furthermore, authentic citations have been provided wherever possible to ensure there is a clear link between the data and categories (Elo & Kyngas, 2007). Dependability, or how stable the findings are over time, and transferability, or how generalizable the findings were to other contexts, were outside the scope of this study.

4.2.3 Mixed Methods Integration

The overall purpose of this study was to evaluate the long-term impact of the McMaster BC by using both qualitative and quantitative methods in a Convergent Parallel Mixed Methods (CPMM) design. Thus, a joint display table was used to integrate the quantitative results from the repeated measures ANOVAs with the qualitative results from the QCA. A joint display table is a mixed methods approach to organizing the quantitative and qualitative results into a single table or graph (Creswell & Plano Clark, 2018). This method was selected as it is commonly used to highlight the similarities and differences between the data in convergent studies (Creswell & Plano Clark, 2018). Themes from the QCA that related to OSCE performance were manually identified and placed in the joint display table. Following this procedure, we examined whether the quantitative and qualitative data were congruent or incongruent. Figure 1 outlines the CPMM process.

4.3 Results

4.3.1 Quantitative Results

Demographics.

Of the 31 residents in the 2015 traditionally trained (TT) cohort, 16 (52%) completed an OSCE at one year into training, and 29 (94%) completed an OSCE at two years into training. Of the 32 residents in the 2016 boot camp (BC) cohort, 32 (100%) completed an OSCE immediately following the BC, 28 (88%) completed an OSCE at one year into training, and 27 (84%) completed an OSCE at two years into training (Table 2).

Statistical analyses.

To evaluate the BC cohort's OSCE performance across time, a one-way repeated measures ANOVA was conducted. To compare OSCE scores between the 2015 TT and 2016 BC cohorts, two-way, repeated measures ANOVAs were conducted. Across the analyses, the Shapiro-Wilk normality tests suggested that scores on some OSCE stations violated the normality assumption ($p < .05$). However, we chose to continue with the repeated measures ANOVAs as the histograms appeared fairly normal on visual inspection and ANOVAs are robust to normality violations (Blanca, Alarcon, Arnau, Bono, & Bendayan, 2017). Checklist scores and global ratings of entrustment were analyzed separately for each OSCE. As both the TT and BC cohorts were fairly homogeneous (Figure 2), no participants were removed as outliers from the analyses.

BC cohort OSCE checklist scores across time.

A one-way, repeated measures ANOVA explored the BC cohorts' OSCE checklist scores across time (post-BC, one year, and two years into training). The sphericity assumption was met for time (p =.22) and station (p =.10). The ANOVA results suggested a main effect of time [F (2,46) = 3.99, p =.02]; station [F (9, 207) = 36.78, p < .01]; and interaction effect of time by station [F (18, 414) =8.80, p < .01] (Table 3a). Post-hoc, pairwise comparisons with a Bonferroni correction for multiple comparisons suggested a statistically significant difference between the post-BC and two year OSCE checklist scores [t (25) = -2.91, p =.01]; no other statistically significant differences were found (Table 3b; Figure 3). Station-specific scores are displayed in Figure 4.

BC cohort OSCE global rating of entrustment across time.

A one-way, repeated measures ANOVA explored the BC cohorts' OSCE entrustment scores across time (post-BC, one year, and two years into training). The sphericity assumption was met for time (p =.26) and station (p =.43). The ANOVA results suggested a statistically significant main effect of time [F (2, 42) = 58.06, p < .01]; station [F (7, 147) =8.46, p < .01]; and interaction effect of time by station [F (14, 294) =5.85, p < .01; Table 3a]. Post-hoc, pairwise comparisons with a Bonferroni correction for multiple comparisons suggested global ratings of entrustment were significantly different between the post-BC and one year OSCEs [t (24) = -12.20, p < .01]; post-BC and two year OSCEs [t (25) = -18.27, $p < .01$]; and one year and two year OSCEs [t (21) = -3.39, $p < .01$] (Table 3b; Figure 5). Station-specific scores are displayed in Figure 6.

BC vs. TT cohort: checklist scores at one year into training.

A two-way, repeated measures ANOVA compared OSCE checklist scores between the TT and BC cohorts at one year into training. The sphericity assumption was violated ($p < .01$), thus the Greenhouse Geisser correction was used. The corrected ANOVA results suggested a statistically significant main effect of group [F (1, 42) = 21.01, $p < .01$]; station [F (5, 223) = 20.43, $p < .01$]; and interaction effect of group by station [F (5, 42) =2.63, $p = 0.02$] (Table 4a; Figure 7). Post-hoc, simple main effects analyses suggested the BC cohort scored significantly higher than the TT cohort on the instrumentation [F (1, 42) = 8.33, $p = .01$], breaking bad news [F (1, 42) = 9.95, $p < .01$], operating room conduct [F (1, 42) = 20.52, $p < .01$], and handover stations [F (1, 42) = 8.75, $p < .01$] (Table 4b; Figure 8).

BC vs. TT cohort: global rating of entrustment at one year into training.

A two-way, repeated measures ANOVA compared OSCE entrustment scores between the TT and BC cohorts at one year into training. The sphericity assumption was borderline violated ($p = .05$), thus the Greenhouse Geisser correction was used. The corrected ANOVA results suggested a statistically significant main effect of group [F (1, 38) = 20.87, $p < .01$] and station [F (5, 204) = 3.46, $p < .01$]. We did not find a statistically significant interaction effect of group by station [F (5, 38) = 2.09, $p = .06$] (Table 4a; Figure 9). However, post-hoc, simple main effects analyses suggested the BC cohort scored significantly higher than the TT cohort on the entrustment component of the suturing [F (1,41) = 13.69, $p < .01$], recognition of the critically ill patient [F (1, 42) = 22.35, $p < .01$], and handover stations [F (1,42) = 4.25, $p = .01$] (Table 4b; Figure 10).

BC vs. TT cohort: checklist scores at two years into training.

A two-way, repeated measures ANOVA compared OSCE checklist scores between the TT and BC cohorts at two years into training. The sphericity assumption was met (p =.46). The ANOVA results suggested a statistically significant main effect of station [F (9, 459) = 42.70, $p < .01$] and interaction effect of group by station [F (9, 51) =10.45, p <.01]. We did not find a statistically significant effect of group [F (1, 51) = 2.57, $p = .11$] (Table 5a; Figure 11). Post-hoc simple main effects analyses suggested the BC cohort scored significantly higher than the TT cohort on the checklist component of the informed consent [F (1,54) = 20.08, $p < .01$], and suturing station [F (1,54) = 14.97, $p < .01$]. However, the TT cohort performed significantly higher than the BC cohort on the checklist component of the pre-operative preparation [F (1,52) = 13.41, $p < .01$] and instrumentation stations [F (1, 54) = 32.43, $p < .01$] (Table 5b; Figure 12).

BC vs. TT cohort: global rating of entrustment at two years into training.

A two-way, repeated measures ANOVA compared OSCE entrustment scores between the TT and BC cohorts at two years into training. The sphericity assumption was violated ($p < .01$), thus the Greenhouse Geisser correction was used. The corrected ANOVA results suggested a statistically significant main effect of group [F (1, 50) = 4.72, p =.03] and station [F (5, 350) = 7.53, $p < .01$], and a statistically significant interaction effect of group by station [F (5, 50) =5.56, $p < .01$] (Table 5a; Figure 13). Post-hoc, simple main effects analyses suggested the BC cohort scored significantly higher than the TT cohort on the entrustment component of the informed consent [$F_{(1,54)} = 17.43$, $p < .01$] and operating room conduct stations [$F_{(1,53)} = 1.02$, $p = .01$] (Table 5b; Figure 14).

Summary of key quantitative findings.

The BC cohort's OSCE performance improved on the checklists and global ratings of entrustment over the two-year follow-up period (Figures 3 and 5, respectively). At one year into training, the BC cohort scored significantly higher than the TT cohort on the checklist (Figure 7) and global rating of entrustment (Figure 9). At two years into training, the BC cohort scored significantly higher than the TT cohort on the global rating of entrustment (Figure 13), although we did not find a statistically significant difference in checklist scores (Figure 11). Post-hoc, simple main effects analyses suggested that the significant differences between cohorts at both one and two years into training were station-specific (Figures 8, 10, 12, 14). This issue will be further discussed in the mixed methods integration section.

4.3.2 Qualitative Results

Residents' perspectives post-boot camp.

Of the 32 residents who participated in the 2016 BC, 32 (100%) participated in the focus group immediately following the BC (Table 2). The 32 focus group participants were randomly divided into four groups (eight residents per group). Across the four focus groups, 100% of residents agreed that the BC was useful in preparing them for residency and should be implemented with future cohorts. We also identified six major themes across the four focus groups. A summary of focus group themes, sub-themes and sample quotations is provided in Table 6.

Gap between medical school and residency.

Residents suggested the ‘gap’ between medical school and residency is due to a number of factors. Residents have a few months off between medical school and residency where they are not in the clinical environment, so they felt generally out of practice. There was also variability at to when residents’ last surgical rotation occurred. Some residents had not been on a surgical service in six to eight months, and thus were worried about skill degradation (quotation 1.2.1). Residents who had switched institutions from medical school to residency expressed concern over learning hospital logistics. As well, participants felt that as residents they were expected to know more and take on more responsibility for patient care than they had as medical students.

Perceived usefulness of the boot camp.

Residents felt the BC was useful in preparing them for residency as it acted as a refresher of their clinical skills and ‘leveled the playing field’. The BC gave everyone an opportunity to practice their surgical skills; thus, how long it had been since their last surgical rotation felt less important (quotation 2.2.1). Following the BC, residents also felt they had a better understanding of the expectations of new residents and their role in the clinical environment. While residents felt they still had a lot to learn, they thought the BC improved their confidence and allowed them to develop an approach to handling different issues they will encounter in the clinical environment. Lastly, residents felt the social aspect of the BC was extremely useful. Residents were able to meet their peers, senior residents, and faculty in a low-stress environment. They felt this social aspect of the BC provided them with a support system and set the tone for maintaining good relationships throughout the remainder of their training program (quotation 2.5.1).

Boot camp format/timing.

Overall, residents felt the two-week BC was a sufficient length. By the end of the two weeks they were getting anxious to return to clinical service, in particular because they felt that upper-year residents who had been covering their clinical responsibilities while they were at the BC were starting to get burnt out. In terms of BC format, residents enjoyed having didactic sessions in the mornings and lab-based sessions in the afternoon during the BC. Sessions with the breakout activities were considered most useful. Residents expressed concern that the BC was too focused on general surgery and wanted more specialty representation.

Suggestions for future iterations.

In accordance with their comments on the BC format, residents made specific suggestions for improving future iterations of the BC. Mainly, residents wanted more specialty-specific content (quotation 4.1.1). They felt reducing the general surgery focus, having more varied cases, and having facilitators from their own specialties would help with this specialty-specific content. Residents felt that some sessions should be removed or shortened to make room for other content (session-specific feedback and sample quotations are provided in Table 7). Namely, residents felt the Advanced Trauma and Life Support (ATLS) course should be moved into the BC as it covered many of the skills taught in BC. In addition to that, residents wanted specific teaching on prepping and draping patients for the operating room, the electronic medical record system, wound healing, how to teach medical students, laparoscopy, and how to run a resuscitation. Residents warned that as changes are being made to future iterations, it would be important to keep the focus on resident needs, rather than hospital administrative needs. They did not want the BC to 'become a dumping ground' for teaching policy and procedures.

OSCE was not reflective of resident performance.

When asked about whether the OSCE was a good measure of BC effectiveness, residents said the OSCE was likely not representative of their improvement in the BC. Residents felt many OSCE stations were unrealistic due to low-fidelity models, acting ability of the mock patients, or the scenarios themselves (quotation 5.1.1). Residents felt the OSCE marks were very subjective, and evaluators often focused on their medical knowledge or emotional response, rather than the task at hand (quotation 5.4.1). Residents also felt that the eight minutes allotted for each station was not enough time. They felt rushed during the OSCE and thought their performance suffered, particularly in the suturing and multiple-choice stations. Lastly, residents felt the multiple-choice question station was not a good way to evaluate what they had learned in the instruments, radiation safety and electrosurgery sessions.

Alternative ways to measure boot camp effectiveness.

Residents suggested that there might be alternative ways to capture BC effectiveness. Residents said there were many opportunities within BC sessions for facilitators to evaluate their performance. Looking at how their performance changed across BC sessions may be more reflective of their skill, rather than a single OSCE station at the end. Residents also indicated that capturing narrative feedback might be another way to explore BC effectiveness, as the numeric scores did not mean very much to them at this stage in training. The feedback they received from facilitators was much more meaningful (quotation 6.2.1). Lastly, residents suggested that we should follow up with them in a few months to see how they felt once they had spent some time working in the clinical environment.

Residents' perspectives at six months into training.

Of the 32 residents who participated in the 2016 BC, 25 (78%) completed the follow-up survey in January 2017 (Table 2). The majority of residents (76%) still felt the BC was definitely useful in preparing them for residency; however, 12% of residents felt it was only somewhat useful, and 12% felt it was not useful. When asked if the BC should be rerun for future cohorts, 84% of residents said yes, 8% did not provide a response, and the other 8% said no. The following themes describe residents' rationale for their ratings. A summary of themes, sub-themes and sample quotations is also provided in Table 8.

Perceived usefulness of the boot camp.

Most residents felt the BC was useful in reviewing expectations for new residents. Residents also appreciated having the opportunity to meet other residents. Residents said these two things significantly decreased their stress when starting residency. Despite these benefits, some residents felt nothing but time and experience can truly prepare someone for residency. Thus, while they thought the BC should be rerun for future cohorts, it may not truly 'prepare them' for residency. Even with a BC, starting clinical service was a very steep learning curve (quotation 20.4.2).

Suggestions for future iterations.

At six months into training, residents felt that the BC would be more useful if it included specialty-specific content. Residents also suggested that BC participants should do some clinical activities during the BC period. They felt being removed from clinical service for a full two weeks was extremely disruptive to other residents and faculty in their specialty, and they missed out on valuable clinical learning opportunities (quotation 21.2.3).

Program Directors' perspectives at six months into training.

Of the nine Program Directors involved in Surgical Foundations at McMaster, 7 (78%) completed a follow-up interview on the BC at six or seven months into training (January – February 2017). Demographics are presented in Table 2. Of the seven participants, five (71%) had participated in the BC as a facilitator. During the interviews, six Program Directors (84%) said the BC was useful in preparing residents for clinical service; however one Program Director (14%) felt the BC was not useful in preparing residents. When asked if the BC should be rerun for future cohorts, six (84%) of the Program Director participants said yes, while the one Program Director who had felt the BC was not useful, said the BC should not be continued. Table 9 displays residents' and Program Directors' perspectives on the BC usefulness over time. In addition to this, we identified four major themes across the Program Director interviews. A summary of themes, sub-themes and sample quotations are provided in Table 10.

Perceived usefulness of the boot camp.

Program Directors felt the BC was useful in easing the transition between medical school and residency, as it seemed to decrease resident stress levels. Program Directors also felt that residents come to residency with different training experiences, so the BC ensured everyone was on the same page and knew what was expected of them. Program Directors felt that disseminating the information in a BC program was more efficient than setting up teaching time for each individual specialty. This was particularly true for the smaller programs that only have one or two incoming residents per year (quotation 22.4.1). Lastly, Program Directors thought the BC provided residents with a valuable opportunity to interact with their peers and get to know people outside of their specialty.

Feedback from the program.

When asked about feedback they had received from residents and faculty in their program, Program Directors said that the residents had some suggestions but overall their feedback was very positive. Program Directors stated that faculty provided little feedback either way. However in one program, faculty had been very negative and did not want residents attending the BC as it disrupted their clinical service (quotation 23.2.2).

Personal observations of the boot camp cohort.

When asked about their personal observations of the residents who participated in the BC, six Program Directors (84%) said they work with first-year residents regularly. Of those that do work with first-year residents, Program Directors found it too difficult to tell if there were any performance differences between the BC cohort and previous TT cohorts (quotation 24.1.1). However, they said there were definitely no negative effects of the BC. In terms of teamwork, most Program Directors felt that teamwork and collegiality between residents was already high, so they did not think the BC had changed that for the first-year residents.

Suggestions for future iterations.

Lastly, Program Directors felt the BC should be continually refined to make the program as efficient and effective as possible. Some programs already have specialty-specific BCs; however, for the programs that do not, Program Directors wanted more specialty-specific content in the BC. Program Directors also expressed interest in adding more non-technical skills to the BC. They stated that while residents may not like it, teaching technical skills (like suturing) in the clinical environment is easy and residents will improve with practice. Teaching non-technical skills is more difficult, and is usually where residents get into trouble (quotation 25.3.1).

4.3.3 Mixed Methods Results

During the mixed methods integration phase (Figure 1), qualitative and quantitative findings were compared and contrasted. As previously mentioned, the two-way, repeated measures ANOVA results suggest the BC cohort scored significantly higher than the TT cohort on both the checklist and entrustment scores at one year into training, and the entrustment scores at two years into training. However, when broken down by OSCE station, the trends are less clear. The BC cohort scored higher than the TT cohort on 6/11 topics at one year into training (suturing, instruments, breaking bad news, operating room conduct, recognition of the critically ill patient, and handover). We did not find significant differences on the other five topics. At two years into training, the BC cohort scored significantly higher than the TT cohort on 3/11 topics (informed consent, suturing, and operating room conduct), while the TT cohort performed significantly higher on 2/11 stations (pre-operative preparation and instruments). We did not find statistically significant differences on the remaining six stations.

By comparing the station-specific scores with the qualitative findings, it seems the BC sessions that residents found particularly useful (informed consent, suturing, handover, breaking bad news, and recognition of the critically ill patient) were often associated with differences in OSCE performance. Notably, residents stated they had enjoyed those sessions as they included breakout sessions for them to repeatedly practice the skill(s) (theme 3.3). In terms of the other stations, qualitative findings suggest that residents felt the pre-operative preparation session included too much information to digest (theme 8.1). Accordingly, the BC cohort performed the same, or worse than the TT cohort on the pre-operative preparation station. The residents felt that the multiple-choice questions were neither relevant nor accurate measures of their knowledge (theme 5.5). The performance results on the multiple-choice questions were mixed; sometimes the BC cohort outperformed the TT, sometimes the TT outperformed the BC cohort, or there were no measurable differences. Residents felt that parts of the operating room conduct session were too long and not relevant (theme 19.1), however they still outperformed the TT cohort on the OSCE station at one and two years into training. Lastly, while there were no differences in the BC and TT cohorts on the hand ties task, there were no comments provided on the session or station (positive or negative). Table 11 presents a summary of the OSCE results by station across the two years, along with themes from the post-BC resident focus group.

In addition to the session-specific feedback, residents felt that OSCEs might not accurately capture their skills in general. In the post-BC focus groups, residents said the OSCE did not feel realistic due to low fidelity of the models, acting ability of the mock patients, lack of clarity around the scenario/instructions (theme 5.1), and time constraints (theme 5.6). Additionally, residents felt the evaluators were subjectively marking (theme 5.2), and often focused on residents' medical knowledge, as opposed to how well they performed a certain task (theme 5.3).

During the six-month follow-up survey and interviews, neither residents nor Program Directors commented on the OSCE specifically. From their own observations, Program Directors felt it was too difficult to identify whether the BC improved residents' clinical skills/performance compared with previous cohorts (theme 24.1). Despite the perceived lack of performance differences between BC and TT cohorts, both Program Directors and residents still felt that the BC was useful. Both residents and Program Directors suggested the BC refreshed residents' basic clinical knowledge/skills and reviewed the expectations of new residents (theme 20.1, theme 22.2), and provided an opportunity for residents to get to know their peers (theme 20.3, theme 22.3). Participants felt that these properties of the BC improved residents' confidence and decreased residents' stress levels (theme 20.2, theme 22.1).

4.4 Discussion

4.4.1 Evaluating Resident Performance Over Time

As discussed in chapter 1.5.2, concerns for patient safety at the beginning of the academic year are a big motivator of BC programs (recall the July Effect). While we were unable to measure whether the BC improved patient outcomes directly (see section 4.1), we were able to capture residents' OSCE performance over the first two years of training. Traditionally, educators have used average performance across all OSCE stations as an indicator of residents' competence (Harden et al., 1975). When viewing the results this way, our study suggests there were performance differences between the BC cohort and TT cohort at both one and two years into training. However, as residents had to perform different clinical skills in each OSCE station, looking at the average scores does not tell us much about where differences in performance exist. When comparing OSCE results by station, our results were consistent with previous literature suggesting BC- trained residents score significantly higher than TT residents on some stations, but not all (Blackmore et al., 2014).

Up to this point, there has not been any literature that has explored why these station-specific differences exist. Our study was the first to use a Convergent Parallel Mixed Methods (CPMM) approach to integrate qualitative data with the quantitative OSCE scores. Our findings suggest that stations in which no differences were observed may be the result of session-specific factors (section 4.3.3). For example, BC participants felt the sessions with breakout group work were most useful. Subsequently, those stations were often where the performance differences existed. We know from chapter 2.3 that having the opportunity to engage in deliberate practice and repeatedly practice a skill while receiving feedback is invaluable for novice learners.

However, there were still some station-specific OSCE results that were not explained by what happened in the BC sessions. For example, residents found parts of the operating room conduct session not relevant, yet they scored significantly higher than the TT cohort at both one and two years into training. It is possible these results were due to rater bias, as OSCE examiners were volunteers from the McMaster Department of Surgery and might have known the cohort to which residents belonged (BC or TT). However, our findings were a product of multiple raters across multiple years. We also attempted to mitigate the effects of rater bias by splitting residents into two identical OSCE tracks. One set of raters evaluated half of the TT and BC cohorts, and a second set of raters evaluated the other half for each OSCE. Additionally, as the TT cohort was one year ahead of the BC cohort, we compared OSCE results by time spent in training rather than OSCE performance on a specific day. For example, in the June 2017 OSCE the TT cohort was two years into training while the BC cohort was only one year into training. Thus, it is unlikely that raters' prior knowledge of the residents impacted OSCE results. In light of the steps we took to avoid rater bias, we suggest that it is possible that station-specific differences, particularly those that conflict with our qualitative data, might be due to the nature of OSCE assessments rather than performance differences amongst the cohorts.

4.4.2 OSCEs Might Not Be Reflective of True Clinical Performance.

The OSCE has been considered the gold standard for assessment in medical education for a long time (section 4.1). However, residents in our study felt the OSCE did not accurately capture their clinical skills. Residents found the quality of the models (suturing), acting ability of the mock patents (breaking bad news), and scenarios themselves (operating room conduct) not realistic. Because of these limitations, they felt that evaluators were often marking their acting ability or emotional response, rather than ability to do a certain task. Their comments bring forth the idea of fidelity.

In medical education, fidelity refers to the extent to which a simulation represents the true clinical scenario (Grierson, 2014). Originally simulations were thought to have high or low fidelity. More recently, researchers have suggested there may be different components of fidelity (Grierson, 2014; Norman, Dore, & Grierson, 2012). One component of fidelity is the degree of physical or structural realism - whether the simulation looks real. The second component of fidelity is the degree of psychological or functional realism - whether the simulation elicits similar behaviour to what is required in the real world. In our study, we used fake skin pads during the suturing station. While they may have 'looked real', residents said the models 'felt different' than real skin. Thus, the residents had to change their suturing technique. Similarly, we used mock patients for many of the non-technical skills stations, such as informed consent and breaking bad news. While mock patients provide high physical fidelity, other factors such as the level of emotion and a lack of personal knowledge about the patient seemed to reduce the functional fidelity. Residents felt they were not able to deliver a genuine response because the situation did not feel real. Unfortunately, there are feasibility limits to how realistic we can make models in simulation. A skin pad will never be the same as suturing in the operating room. No matter how much training mock patients undergo, if residents do not treat the simulation as 'real', assessing non-technical skills in simulation may have limited value. In addition to the fidelity issues, residents stated evaluator ratings were fairly subjective and the numeric scores were not very useful to residents. Residents found the narrative feedback more useful. We know that feedback is critical to novice skill acquisition (chapter 2.3.7). Given the concerns around the psychological fidelity of OSCEs, it is possible that statistically significant differences in OSCE scores do not translate into significant differences in clinical performance between the two cohorts. These findings suggest that we may get more information from comparing narrative comments rather than the numeric scores.

4.4.3 OSCEs Might Not Be a Good Measure of Boot Camp Effectiveness

Even if OSCE scores do reflect true clinical performance, both residents and Program Directors identified a number of benefits to BCs that were not captured by the OSCE. These benefits included having an opportunity to become acquainted with hospital logistics, to review basic clinical skills and expectations of residents, and to meet other residents. Residents suggested that BCs might not fully prepare them for residency, but the BC provided them with an approach, confidence, and a support system to handle clinical situations. We know from chapter 2.2.5 that reducing extraneous information is particularly important for novice learners, as learning new skills requires a lot of cognitive capacity. Previously, Sonnadara et al. (2012) had suggested that BCs prime learners for future clinical learning, as less time and attention need to be dedicated to the basic information. We believe our qualitative findings, plus the fact that residents' scores improved over the two years, provide evidence to support this claim. We also know from chapter 1.4 that novice learners are under a large amount of stress at the onset of residency, and burnout and suicide rates are a concern. Our results highlight the value of the social component of BCs and creating a support system for novice trainees. In summary, we propose there are more benefits to BCs than direct improvements to resident skills. Thus, even if OSCEs can capture true clinical performance, we suggest that educators should not rely on OSCE scores alone when judging the effectiveness of a BC program.

4.5 Limitations

When the TT cohort started residency, OSCEs were not yet a regular part of the SF curriculum. The program did not mandate annual OSCEs until July 2016. Thus, we were unable to determine whether there were differences between the TT and BC cohorts prior to training. Moreover, as only 52% of the TT cohort completed an OSCE in June 2016, the OSCE results may not be reflective of the entire TT cohort at one year into training. In addition to these limitations, there were many factors out of our control in this study. With the exception of the BC, the SF curriculum was kept the same for the 2015 and 2016 cohorts. However, it is difficult to quantify how much teaching occurs in the clinical environment during the first two weeks of residency. It is possible one cohort received more teaching than the other. During the BC itself, facilitators were volunteers, thus we were unable to control for their facilitation style and whether they followed the session objectives. We were also unable to control for resident study habits over the two year period. While we are unable to identify whether these factors influenced the results, they are highly reflective of medical training contexts; thus we propose these limitations added ecological validity to our study.

4.6 Conclusions

To date, studies have focused on the short-term impact of BC programs using quantitative measures of resident performance. This study was the first to explore BC effectiveness beyond seven months into training. This study was also the first to use a mixed methods approach and integrate quantitative and qualitative data to study BC effectiveness. The quantitative results suggested that on average, the BC cohort scored higher on the OSCE than the TT cohort at both one and two years into training. However, consistent with previous literature, the post-hoc analyses suggest the differences between the BC and TT cohort were station-specific. The qualitative results suggest this may be due to the BC sessions themselves, as residents perceived some sessions as more useful than others. The qualitative results also suggest that OSCEs might not be an optimal measure for capturing BC effectiveness. Residents felt that the OSCE was not reflective of their improvement in the BC. Additionally, both residents and Program Directors felt that there are many benefits to BCs beyond those related to skill improvement. Taken together, using OSCE data alone might not capture the true impact of a BC program. Future research should continue to explore alternative ways to capture BC effectiveness and how those relate to OSCE performance.

In summary, using a mixed methods approach to integrate qualitative and quantitative data offered new insights into the effectiveness of BC training courses and highlighted several contextual factors that contribute to that effectiveness. We hope that through sharing these findings, educators will be provided with data to support implementing BC programs for incoming residents. We also hope that educators build on these findings to continue exploring the most effective ways to teach novice residents.

Table 1

OSCE Schedule

OSCE	**2015 Traditionally trained (TT) Cohort**	**2016 Boot Camp (BC) Cohort**
OSCE #1 Post-Boot Camp	N/A	July 2016
OSCE #2 End of 1st Year	June 2016	June 2017
OSCE #3 End of 2nd Year	June 2017	June 2018

Table 2

Participant Demographics

Data Collection	Participants			Gender (% Female)	Gender (% Male)
	2015 TT Cohort	2016 BC Cohort	Surgical Program Directors		
OSCE #1 Post-Boot Camp	N/A	32/32 (100%)	N/A	14/32 (44%)	18/32 (56%)
OSCE #2 End of 1st Year	16/31 (52%)	28/32 (88%)	N/A	19/44 (43%)	25/44 (57%)
OSCE #3 End of 2nd Year	29/31 (94%)	27/32 (84%)	N/A	32/56 (57%)	24/56 (43%)
Post-BC Resident Focus Group	N/A	32/32 (100%)	N/A	14/32 (44%)	18/32 (56%)
Six Month Resident Survey	N/A	25/32 (78%)	N/A	Not available	Not available
Six Month Program Director Interviews	N/A	N/A	7/9 (78%)	2/7 (29%)	5/7 (71%)

Note: The six month surveys were anonymous. Thus, we cannot comment on the gender proportions.

Table 3a

OSCE Scores for the 2016 BC Cohort Across Time: ANOVA Results

OSCE Component	**Comparison**	***df***	***F-value***	***p***
Checklist Scores	Time	2, 46	3.99	**.02**
	Station	9, 207	36.78	**< .01**
	Time x Station	18, 414	8.80	**< .01**
Global Ratings of Entrustment	Time	2, 42	58.06	**< .01**
	Station	7, 147	8.46	**< .01**
	Time x Station	14, 294	5.85	**< .01**

Note: Statistical significance was considered when $p < .05$

Table 3b

OSCE Scores for the 2016 BC Cohort Across Time: ANOVA Post-Hoc Comparisons

OSCE Component	**Comparison**	***df***	***t-value***	***p***
Checklist Scores	Post-BC vs. Year 1	27	- 0.10	1.0
	Post-BC vs. Year 2	25	- 2.91	**.01**
	Year 1 vs. Year 2	23	- 2.26	.10
Global Ratings of Entrustment	Post-BC vs. Year 1	24	- 12.20	**< .01**
	Post-BC vs. Year 2	25	- 18.27	**< .01**
	Year 1 vs. Year 2	21	- 3.39	**< .01**

Note: Statistical significance was considered when $p < .05$

Table 4a

OSCE scores Between the 2015 Traditionally Trained (TT) Cohort and 2016 Boot Camp (BC) Cohort at One Year into Training: ANOVA Results

OSCE Component	**Comparison**	***df***	***F-value***	***p***
Checklist Scores	Group	1, 42	21.01	**< .01**
	Station	5, 223	20.43	**< .01**
	Group x Station	1, 42	2.63	**.02**
Global Ratings of Entrustment	Group	1, 38	20.87	**< .01**
	Station	5, 204	3.46	**< .01**
	Group x Station	5, 38	2.09	.06

Note: Statistical significance was considered when $p < .05$

Table 4b

OSCE scores Between the 2015 Traditionally Trained (TT) Cohort and 2016 Boot Camp (BC) Cohort at One Year into Training: ANOVA Post-Hoc Comparisons

OSCE Component	**Comparison**	***df***	***F-value***	***p***
Checklist Scores	Informed consent	1, 42	0.45	.51
	Suturing	1, 42	1.46	.23
	Pre-operative preparation	1, 42	0.48	.49
	Instrumentation	1, 42	8.33	**.01**
	Radiation safety	1, 42	1.41	.24
	Electrosurgery	1, 42	0.17	.68
	Breaking bad news	1, 42	9.95	**< .01**
	Operating room conduct	1, 42	20.51	**< .01**
	Recognition of the critically ill patient	1, 42	6.20	.02
	Handover	1, 42	8.75	**< .01**
Entrustment Scores	Informed Consent	1, 42	0.21	.65
	Suturing	1, 41	13.69	**< .01**
	Pre-operative preparation	1, 42	4.10	.05
	Breaking bad news	1, 41	6.11	.02
	Operating room conduct	1, 41	0.87	.35
	Recognition of the critically ill patient	1, 42	22.35	**< .01**
	Handover	1, 42	6.61	**.01**
	Hand Ties	1, 40	4.25	.05

Note: Statistical significance was considered when $p < .01$

Table 5a

OSCE scores Between the 2015 Traditionally Trained (TT) Cohort and 2016 Boot Camp (BC) Cohort at Two Years into Training: ANOVA Results

OSCE Component	**Comparison**	***df***	***F***	***p***
Checklist Scores	Group	1, 51	2.57	.11
	Station	9, 459	42.70	**< .01**
	Group x Station	9, 51	10.45	**< .01**
Global Ratings of Entrustment	Group	1, 50	4.72	**.03**
	Station	5, 350	7.53	**< .01**
	Group x Station	5, 50	5.56	**< .01**

Note: Statistical significance was considered when $p < .05$

Table 5b

OSCE scores Between the 2015 Traditionally Trained (TT) Cohort and 2016 Boot Camp (BC) Cohort at Two Years into Training: ANOVA Post-Hoc Comparisons

OSCE Component	**Comparison**	***df***	***t-value***	***p***
Checklist Scores	Informed consent	1, 54	20.08	**< .01**
	Suturing	1, 54	14.97	**< .01**
	Pre-operative preparation	1, 52	13.41	**< .01**
	Instrumentation	1, 54	32.43	**< .01**
	Radiation safety	1, 54	2.51	.12
	Electrosurgery	1, 54	6.18	.02
	Breaking bad news	1, 54	3.67	.06
	Operating room conduct	1, 54	0.00	.97
	Recognition of the critically ill patient	1, 54	1.08	.30
	Handover	1, 53	3.91	.05
Global Ratings of Entrustment	Informed Consent	1, 54	17.43	**< .01**
	Suturing	1, 54	4.70	.03
	Pre-operative preparation	1, 53	0.53	.47
	Breaking bad news	1, 54	5.47	.02
	Operating room conduct	1, 53	7.40	**.01**
	Recognition of the critically ill patient	1, 54	6.04	.02
	Handover	1, 53	1.02	.32
	Hand Ties	1, 52	0.18	.68

Note: Statistical significance was considered when $p < .01$

Table 6

Post-Boot Camp Resident Focus Group (Overall Themes)

Theme	Sub-theme	Example Quotations
'Gap' between medical school and residency occurs for a number of reasons (1)	Time off between medical school and residency (1.1)	*"Its been quite a while since [clerkship], and you do lose [skills] pretty fast. And I just feel that if we were to be thrown full blown into residency [it] would be a hard thing to do. I am not the same person I was in clerkship and I am probably a little worse." – Resident focus group #2 (1.1.1)*
	Time since last surgical rotation (1.2)	*"Even though we had done a lot of electives [in surgery], we had a couple months off after medical school, before that we had a couple of months to prepare for the exam, and before that we were kind of done clerkship. So I hadn't sutured in, for instance, like 6-8 months; I was worried about being kind of rusty." – Resident focus group #2 (1.2.1)*
	Understanding hospital logistics (1.3)	*"The thing I felt least prepared for was actually like working in the system and understanding, like the EMR and what's on the computer. Things are done so differently from the hospital I am used to, and when I did an elective here, I didn't actually have a chance to like work in the settings that I am working in now." – Resident focus group #2 (1.3.1)*
	Different expectations/responsibilities between clerks and residents (1.4)	*"I think medical school focuses on the knowledge base, identifying certain patterns, whereas residency obviously is more service-based. You are actually supposed to do something about it." – Resident focus group #1 (1.4.1)*
Perceived usefulness of the boot camp (2)	100% of residents agreed the boot camp should be run for future cohorts (2.1)	*"Yes [run a boot camp again], especially next year where things are competency based…. I think boot camp might be a good way of jump-starting those basic competencies." – Resident focus group #2 (2.1.1)*
	Boot camp acted as a refresher/leveled playing field (2.2)	*"I think [boot camp is] definitely nice way to level the playing field per say. Everyone came from different backgrounds, everyone learnt different things, different ways and have like different skill levels, so kind of having two weeks to at least make sure everyone is up to the same baseline level is like "you have all done this", "you have all learnt this skill", then it's easier for staff to know that they can trust that this resident has gone through boot camp and is like able to do some things." – Resident focus group #2 (2.2.1)*
	Boot camp allowed residents to develop an *approach* to handling situations in the clinical environment (2.3)	*"I think [the boot camp] just kind of re-iterated that we can handle the common issues on call. Like we were saying earlier, for some of us it has been a lot of months since we've been in the OR, so just kind of re-acquainting ourselves with the technical skills." – Resident focus group #4 (2.3.1)*

Table 6 Cont'd

Theme	Sub-theme	Example Quotations
Perceived usefulness of the boot camp (2)	Boot camp reviewed the expectations of residents (2.4)	*"Now I know what is expected of me when I am starting like residency, so [boot camp] was really helpful for me to say yes I can do this and I am confident enough to do it on my own." – Resident focus group 4 (2.4.1)*
	Social aspect of boot camp was major benefit (2.5)	*"It's been really nice to meet people from other specialties that you'll be seeing or consulting with over the next 5 years while you're here. It's like, oh, I've met you, we've chatted, you know, and then that feels better in terms of building collegiality amongst programs going forward." – Resident focus group 1 (2.5.1)*
Comments on the boot camp format/timing (3)	Two week period could be condensed /refined (3.1)	*"[The boot camp] could have been more concise... Having two weeks of like transition time is an appropriate amount of time, but I think some of the sessions were for example too long and we didn't really need to have as much time those specific content areas." – Resident focus group 2 (3.1.1)*
	Format (didactic mornings, lab afternoons) was well received (3.2)	*"I kind of liked having the lectures in the morning, I found I paid more attention. In the afternoons it is nice having something else." – Resident focus group #4 (3.2.1*
	Sessions with breakout sessions were most effective (3.3)	*"The sessions where you go off do your own thing, get feedback from each other, and then come back for discussion were the best…It was practical." – Resident focus group 3 (3.3.1)*
	Residents felt that it was useful to engage in observational learning (3.4)	*"You got to see what works and what doesn't work, try to change how you do things. If something went well then you try to do it like that, and if it didn't work well, you try to figure out why it didn't work well." – Resident focus group 4 (3.4.1)*
	Boot camp content was too general surgery focused (3.5)	*"Almost all of the scenarios yesterday were very geared towards general surgery, and a lot of people were not in general surgery (agreement from group)…" –Resident focus group 3 (3.5.1)*
	Residents seemed anxious to get to work by the end of boot camp (3.6)	*"The two weeks before kind of felt like it just delays the inevitable. You can do all the sim that you want, and sims are great but you still need to put it into practice and until you practice you're not going to feel comfortable." Resident focus group 1 (3.6.1)*
	Residents felt pressure from their programs to be focusing on specialty specific content/to get back on clinical service (3.7)	*"Talking to the other juniors on my team who are not in boot camp who are having to cover call for those of us who are in boot camp, they have been doing 1 in 2 call for the last few weeks and they are feeling quite burned out because of it." – Resident focus group 4 (3.7.2)*

Table 6 Cont'd

Theme	Sub-theme	Example Quotations
Suggestions for future iterations (4)	Residents wanted more specialty specific content (4.1)	*"I know this is sort of supposed to be back to the basics because there will be specialty specific stuff as we go through [training], but there is a lot of basic specialty specific stuff. Like I bet if you asked the staff they would say it would be REALLY helpful if residents knew how to do this before their first time." – Resident focus group 3 (4.1.1)*
	Residents felt ATLS should be included in boot camp (4.2)	*"Somebody said that some of the technical skills workshops were quite redundant… I thought that if we could include certification courses, like ATLS, in the boot camp that would save time since ATLS has central lines, tubes, etc." – Resident focus group 3 (4.2.1)*
	Having mock patients complete more training would increase authenticity (4.3)	*"I mean in OSCEs it might have been helpful to have the formally trained standardized patients who are able to really give you those emotions. I found with my volunteer patient you know there was no anger, there was no real being upset, you didn't have to deal with the emotions or anything like that." – Resident focus group 4 (4.3.1)*
	Including more varied cases would increase usefulness of practice scenarios (4.4)	*"Could be less morbid stuff too. The three cases to practice were they were dying…But bad news can be a bad outcome from your surgery, or that you need to have this X operation."– Resident focus group 3 (4.4.1)*
	Residents wanted faculty and residents from their own specialties to facilitate boot camp session (4.5)	*"And the staff and residents who are there, if there was someone from every specialty, that also would be beneficial." – Resident focus group 4 (4.5.1)*
	Residents suggested additional sessions to add to the boot camp, including: draping, EMR, teaching medical students, wound healing, laparoscopy, & running a code (4.6)	*"I heard someone saying it would have been nice to have a formal, maybe 45 minute, session with the residents on going through meditech and citrix. Not exactly on how to use it, but this is where you find vitals, this is how you access the flow chart, you have to hit period enter. Stuff that is not intuitive to do." – Resident focus group 3 (4.6.1)*
	Residents warned that organizers should keep the boot camp focused on resident needs rather than hospital administrative needs (4.7)	*"As this course goes on every regulatory body in the hospital will find out about it and try to get a session about some policy procedure that they feel all surgical persons need to know. If you guys aren't careful to keep this designed by surgeons for surgeons, in three years the morning session may be HIPA and privacy and then the afternoon about infection control, and then it will be the scrub policy…I am not saying that's not important, just if you are not careful it will become a two week policy and procedures rundown." – Resident focus group 2 (4.7.1)*

Table 6 Cont'd

Theme	Sub-theme	Example Quotations
Residents felt the OSCE was not reflective of their improvement (5)	OSCE did not feel realistic to the residents (models, mock patients, scenarios) (5.1)	*"In an exam situation you're like okay, [the mock patient] looks very healthy and I'm supposed to tell them that they have this fake disease, and they're not responding in a way a person would in a normal situation…I think it doesn't look genuine because it's not because the whole thing is contrived."* *– Resident focus group 1 (5.1.1)*
	Residents felt that evaluator marks were subjective (5.2)	*"I think part of the issue is that there is a huge amount of subjectivity [on OSCE evaluations]." – Resident focus group 4 (5.2.1)*
	Some stations felt geared towards medical knowledge rather than the task at hand (5.3)	*"[Informed consent] focused a lot more on knowledge than the consent component…if you are going to pick a procedure that we don't know, that's fine, but tell us what it entails." – Resident focus group 2 (5.3.1)*
	Residents felt that some stations measured acting ability rather than competence at the task (e.g. breaking bad news) (5.4)	*"The [OSCE] feedback is that you are not showing you are emotional enough, very valuable for sure, and maybe sometimes it has to be contrived in the real scenario, you don't know the patient and you don't have that emotional connection. But it's tough to kind of put together [that emotion] especially in an OSCE situation, like [I was] suturing, [next] I took consent, now boom you have to tell someone they have cancer." – Resident focus group 3 (5.4.1)*
	Residents felt the multiple-choice questions were not relevant/fair (5.5)	*"To be honest, I hated the multiple choice questions on electrosurgery…they were absurd. But, I liked the lecture itself, it taught me some stuff that I did not know." – Resident focus group 2 (5.5.1)*
	Residents felt that time constraints affected their performance (5.6)	*"I felt that the OSCE station didn't reflect at all what my improvement has been…the only thing that anyone cared about was that 8 minutes when you were trying to rush through. I feel like there were so many opportunities during the two weeks to have someone come by and see you actually suturing pig skin…." – Resident focus group 1 (5.6.1)*
Alternative ways to measure boot camp effectiveness (6)	Daily evaluations (6.1)	*"I think it would be nice if we could have sort of a checklist for every session we are doing, instead of objectives. Have a checklist or evaluation, like we know how to do this, we checked our safety, we know how to use the proper blade or instruments and stuff like that." – Resident focus group 4 (6.1.1)*
	Narrative Feedback (6.2)	*"This is only 2 weeks into a 5-year residency. The numeric scores that I got on anything right now are not as valuable to me as the feedback getting from staff surgeons." – Resident focus group 2 (6.2.1)*
	Follow up in a few months (6.3)	*"A couple of blocks from now, ask what was useful, what did you take away and what was completely useless." – Resident focus group #1 (6.3.1)*

Table 7

Post-Boot Camp Resident Focus Group (Session Specific Feedback)

Boot Camp Session	Feedback	Sample Quotations
Introduction to instruments (7)	Useful but could have used more specialty specific content (7.1)	*"I think [instruments] was a good session. It was helpful. Even if you knew what most of the instruments were, it was helpful to go back over them and remember what the things were called." – Resident focus group #2 (7.1.1)*
Pre-operative preparation (8)	Too much information, too long (8.1)	*"The anesthesia one was far too much information crammed into like an hour and was not really directed at what we would need to know or do." – Resident focus group #2 (8.1.1)*
Suturing (9)	Good sessions but the suturing materials/models were low quality (9.1)	*"[Suturing was] really good. I noticed significant improvement [over the boot camp]." – Resident focus group #1 (9.1.1)* *"Just some side feedback, on the suturing sessions, people were complaining about the pigskin being too tough, and for the subcuticular, it would have been great to have smaller needles with an actual monochrome appropriate for subcuticular." –Resident focus group #3 (9.1.2)*
Critical incident management (10)	Important topics but does not need to take place in the boot camp (10.1)	*"I just felt like that was a lot of details, like what exactly procedure will happen if you mess up and the board reviews you…I mean is helpful to know, but like maybe not in boot camp." – Resident focus group #2 (10.1.1)*
Recognition of the critically ill patient (11)	Important topic, some redundancy between the two lectures (11.1)	*"Recognition of the critically ill patient was good. I just thought that there was a lot of overlap between the two presenters." – Resident focus group #2 (11.1.1)*
Tubes/Drains (12)	Redundant with ATLS (12.1)	*"I think [tubes and drains] was good, but ATLS covered that." – Resident focus group #2 (12.1.1)*
Radiation Safety (13)	Excellent facilitator but topic was not relevant to everyone (13.1)	*"I don't know how relevant [radiation safety] was, but it was interesting. I like the way that [the facilitator] engaged with the topic." – Resident focus group #1 (13.1.1)*
Electrosurgery (14)	Too long; should be condensed and focus on practical skills (14.1)	*"I think it is important to understand how [cautery] works and why it works, but I found it a little too detailed for my lack of physics knowledge." – Resident focus group #4 (14.1.1)*

Table 7 Cont'd		
Boot Camp Session	**Feedback**	**Sample Quotations**
Ward management (15)	Residents had expected something different (15.1)	*"I saw the topic of Ward Management and was like, okay great he's going to give us tips on if someone's tachycardic or something…And then none of that, I don't think, was reviewed… [that session was a] missed opportunity" – Resident focus group #1 (15.1.1)* *"I think replacing the ward management content with something like common calls to the ward, like that would be very helpful…" – Resident focus group #2 (15.1.2)*
Informed Consent (16)	Useful session, specifically completing forms (16.1)	*"Just going through and filling out the form before you have to do it in the hospital is useful." – Resident focus group #4 (16.1.1)*
Breaking bad news (17)	Some residents felt the session was useful; others felt it had been covered in medical school (17.1)	*"The lecture we had that morning was something that we had all heard in medical school, it's all been done like again and again and again, and we know the SPIKEs mnemonic, that's not the area of weakness, it's one of those things you will only get better at by practicing it a bunch of times, practicing with real patients, practicing it we all know the concepts behind it." – Resident focus group #3 (17.1.1)*
Handover (18)	Resident felt handover breakout activity was very useful, didactic less so. Session morphed into discussion on how to supervise/teach medical students (18.1)	*"I think it was super useful, like what Participant X was saying [earlier] about how to read the charts, I thought it was a good exercise." – Resident focus group #4 (18.1.1)* *"I just found the lectures surrounding it, not as useful, like I don't know, maybe that was just me, and maybe like I said having specialty specific prompts would have probably made it better" – Resident focus group #4 (18.1.2)*
Operating room Conduct (19)	Residents felt it was too long and unnecessary (19.1)	*"Not practical. We all know how to dress and act in the OR. We learned that in med school…" – Resident focus group #1 (19.1.1)*

Table 8

Six Month Resident Follow-Up Survey

Theme	Sub-theme	Sample Quotations
Perceived usefulness of the boot camp (20)	Boot camp reviewed the expectations of residents (20.1)	*"Yes [boot camp was useful], baseline expectations were set and the gradual transition was useful (mentally as well)." – Resident survey P10 (20.1.1)* *"Good intro to things, set expectations for hospital systems/call." – Resident survey P16 (20.1.2)*
	Decreased stress of starting residency (20.2)	*"Was helpful addressing some worries & some basics of surgical specialties. It helped overcome the first hump of residency." – Resident survey P1 (20.2.1)* *"Apart from educational component, it relieves any anxiety prior to starting" – Resident survey P2 (20.2.2)*
	Social aspect of boot camp was major benefit (20.3)	*"It was a good way to get to know my co-residents…" – Resident survey P18 (20.3.1)* *"Nice way to transition. Also good way to know the other surgical residents." – Resident survey P17 (20.3.2)*
	Nothing but time and experience can truly prepare someone for residency (20.4)	*"Yes, I think it was a valuable bridge, although it is still a steep learning curve once our rotations start." – Resident survey P21 (20.4.1)* *"I don't think any didactic/procedural sessions can prepare you for transition to residency. Time & experience is the only thing that makes you feel more comfortable." – Resident survey P22 (20.4.2)*
Suggestions for future iterations (21)	Inclusion of more specialty specific content (21.1)	*"Include more specialty specific technical skills." –Resident survey P3 (21.1.1)* *"[Include] more specialty specific sessions." – Resident survey P19 (21.1.2)*
	Residents should participate in some clinical activity/call during the boot camp to not miss learning opportunities (21.2)	*"Yes [boot camp should be], if some time on service is included into boot camp." – Resident survey P25 (21.2.1)* *"[Residents] should participate in call throughout [boot camp] to not miss too much clinical experience." – Resident survey P3 (21.2.2)* *"This boot camp took away 2 weeks from my own service in which I only get 3 months of this year." – Resident survey P18 (21.2.3)*

Table 9

Perceived Boot Camp Usefulness Across Time

Data Collection	Was boot camp useful in preparing new residents for residency?				Should the boot camp be run for future cohorts of incoming residents?			
	Yes	Somewhat	No	No response	Yes	Maybe	No	No response
Resident focus group (Post-BC)	32/32 (100%)	0 (0%)	0 (0%)	0 (0%)	32/32 (100%)	0 (0%)	0 (0%)	0 (0%)
Resident follow up survey (6-months)	19/25 (76%)	3/25 (12%)	3/25 (12%)	0 (0%)	21/25 (84%)	0 (0%)	2/25 (8%)	2/25 (8%)
Program Director follow-up interviews (6-months)	6/7 (86%)	0 (0%)	1/7 (14%)	0 (0%)	6/7 (86%)	0 (0%)	1/7 (14%)	0 (0%)

Table 10

Six Month Program Director Interviews

Theme	Sub-theme	Sample Quotations
Perceived usefulness of the boot camp (22)	Decreased resident stress-levels (22.1)	*"I think people like the concept of easing into something at the beginning of their residency. I think people are nervous, I think they are anxious, I think they are now actually forced to make some decisions, take some responsibility, sign off on their orders, they are doctors per say now. I think that makes people nervous so it is nice to have two weeks where you kind of ease in and get into it. Although, my personal bias is, they're adults, get on with it." – Program Director interview P6 (22.1.1)* *"[The residents] are probably less stressed when they go on their service, probably less stressed, because I think they're much better prepared, so I think it is a good thing." – Program Director interview P7 (22.1.2)*
	Ensured everyone is on the same page when they start (22.2)	*"I do think [the boot camp] gave [residents] at least a starting point to work with. Residents can come in at all different experience levels…so I do think [the boot camp] helped them get ready for the clinical experience just by leveling the playing field and knowing what the expectations were." – Program Director interview P1 (22.2.1)*
	Provided residents with an opportunity to interact with their peers (22.3)	*"With all of the residents together. So I think it was probably a good set up, those first two weeks, that's you know the first day you come in and meet all the other residents that are on the different programs or even in your own program, they get to mingle together, get to know each other, so all the social side of it also. And I think it is very good." – Program Director interview P7 (22.3.1)*
	Disseminated information in an efficient manner (22.4)	*"[The boot camp] definitely is appreciated from my end because it saves me some stuff too for our residents in terms of having to give these sessions, especially if you [don't have many residents in the program], it is not always worthwhile session to sit one person down. So I think it has made it definitely more efficient in how we are transmitting some really, really useful information that may have not been transmitted in a formal process before." – Program Director interview P1 (22.4.1)*
Feedback from program (23)	Received positive feedback from residents (23.1)	*"I thought overall it was good. The residents seemed to be very positive about their experience." – Program Director interview P4 (23.1.1)* *"I can't remember the specifics. I think overall they quite liked it. So the feedback was overall positive." – Program Director interview P6 (23.1.2)*

Table 10 Cont'd

Theme	Sub-theme	Sample Quotations
Feedback from program (23)	Received no feedback or negative feedback from faculty due to disruptions to clinical service (23.2)	*"Haven't heard outside staff truly commenting on it." – Program Director interview P1 (23.2.1)* *"The faculty were less than enthralled; the faculty who didn't participate were less than enthralled because it impacted clinical service and have actually suggested we should quit boot camps but we told them that is not going to happen." – Program Director interview P3 (23.2.2)*
Personal observations of boot camp cohort (24)	Too difficult to tell if there are performance differences between the BC cohort and previous cohorts (24.1)	*"It might be a little bit too hard to say [if there is a performance difference in the boot camp cohort], and it's hard to, even if I do see one, is it because of the foundations, or is a better/different batch of learners, or what is it right?" – Program Director interview P2 (24.1.1)*
	Felt teamwork and collegiality was already high within program, so boot camp would not have changed that (24.2)	*"I think, in our program [the residents] already work extremely well together, we have a extremely supportive and dependent program, so I haven't noticed that it is different [with the boot camp cohort]." – Program Director interview P3 (24.2.1)*
	No negative effects of the boot camp (24.3)	*"No negative effects of the boot camp by all means, no." – Program Director interview P1 (24.3.1)*
Suggestions for future iterations (25)	Keep refining the boot camp content based on feedback (25.1)	*"Continuing to refine the content so that we, you know, we didn't get everything right this year and I know we are working on refining the content, to take that feedback, then I think it's going to become more valuable as time goes on and we get the content better." – Program Director interview P3 (25.1.1)*
	Include more specialty specific content (25.2)	*"Probably incorporating some more time, some specialty specific time for learners to have exposure to their specific area." – Program Director interview P5 (25.2.1)* *"I think there needs to be specific sub-specialty half-day at least or something, where we can come together and um show the new residents the specific tools, instrumentation, and general foundational things for their specific areas of practice." – Program Director interview P6 (25.2.2)*
	Focus on teaching the non-technical skills that are difficult to teach in the clinical environment (25.3)	*"The suturing and all that, they are going to learn, the sessions they like, the hands on ones, but those are the ones they are going to learn no problem, it's all the other stuff that is hard for us to teach. The communication ones, the process ones, the medical legal stuff, what have you, that I think I would appreciate as a Program Director [including in boot camp], although I know the residents might not know the value of them right away, I think they would be helpful in the end." – Program Director interview P1 (25.3.1)*

Table 11

Joint Display Table – OSCE Scores and Qualitative Feedback by Station

Station	OSCE Scores @ 1 year				OSCE Scores @ 2 years				Relevant qualitative findings (Theme number)
	Checklist		Entrustment		Checklist		Entrustment		
	Means (Tt, BC)	Sig diff	Means (Tt, BC)	Sig diff	Means (Tt, BC)	Sig diff	Means (Tt, BC)	Sig diff	
Informed consent	68.75,72.96		3.19, 3.29		73.01, 91.48	+	3.38, 4.19	+	Residents felt practicing with mock patients & completing consent forms was useful (16.1); however OSCE felt focused on medical knowledge rather than informed consent (5.3)
Suturing	68.75, 78.57		3.33, 4.25	+	79.01, 98.08	+	4.27, 4.75		Residents felt the sessions were very valuable (9.1); however OSCE model was poor (5.1) and time restraints challenged suturing (5.6)
Pre - operative preparation	96.25, 94.34		3.44, 3.79		92.59, 80.90	-	3.81, 3.65		Residents felt the session included too much information (8.1)
Instrument MCQ	68.75, 82.50	+			85.19, 66.92	-			Residents liked this session (7.1) but felt it was difficult to identify the details of the instruments in the MCQ pictures (5.5)
Radiation safety MCQ	66.87, 72.14				68.52, 75.77				Residents felt the topic may not have been relevant but the speaker was excellent (13.1); no comments were provided on the OSCE station (5.5)
Electro-surgery MCQ	54.69, 52.50				43.11, 53.27				Residents felt MCQ were not relevant or practical (5.5)

Table 11 Cont'd

Breaking bad news	66.67, 81.78	+	3.25, 3.85		79.26, 86.63		3.77, 4.19		Some residents felt the session was useful; others felt it had been covered in medical school (17.1); all felt the OSCE station was unrealistic (5.1, 5.4)
Operating room conduct	56.25, 81.07	+	3.19, 3.44		87.22, 87.41		3.48, 4.19	+	Residents felt parts of the session were not practical (19.1) and the OSCE station was unclear (5.1)
Recog. of the crit. ill patient	79.37, 88.57		2.94, 3.93	+	88.52, 84.04		3.88, 3.35		Residents felt the session was very useful (11.1); no comments were provided on the OSCE station
Handover	80.36, 90.81	+	3.62, 4.21	+	95.65, 87.36		4.15, 4.08		Residents felt the session was very useful (18.1). The only comment on the OSCE station was that it did not take long
Hand Ties			3.81, 3.75				3.96, 4.04		No comments were provided on this topic

Note:
+ Indicates OSCE stations where the BC Cohort scored significantly higher than the TT cohort;
- Indicates OSCE stations where the TT cohort scored significantly higher than the BC cohort;
Blank boxes indicate no statistically significant difference found between the cohorts;
Gray boxes indicate that the checklist or entrustment rating was not collected for that station.

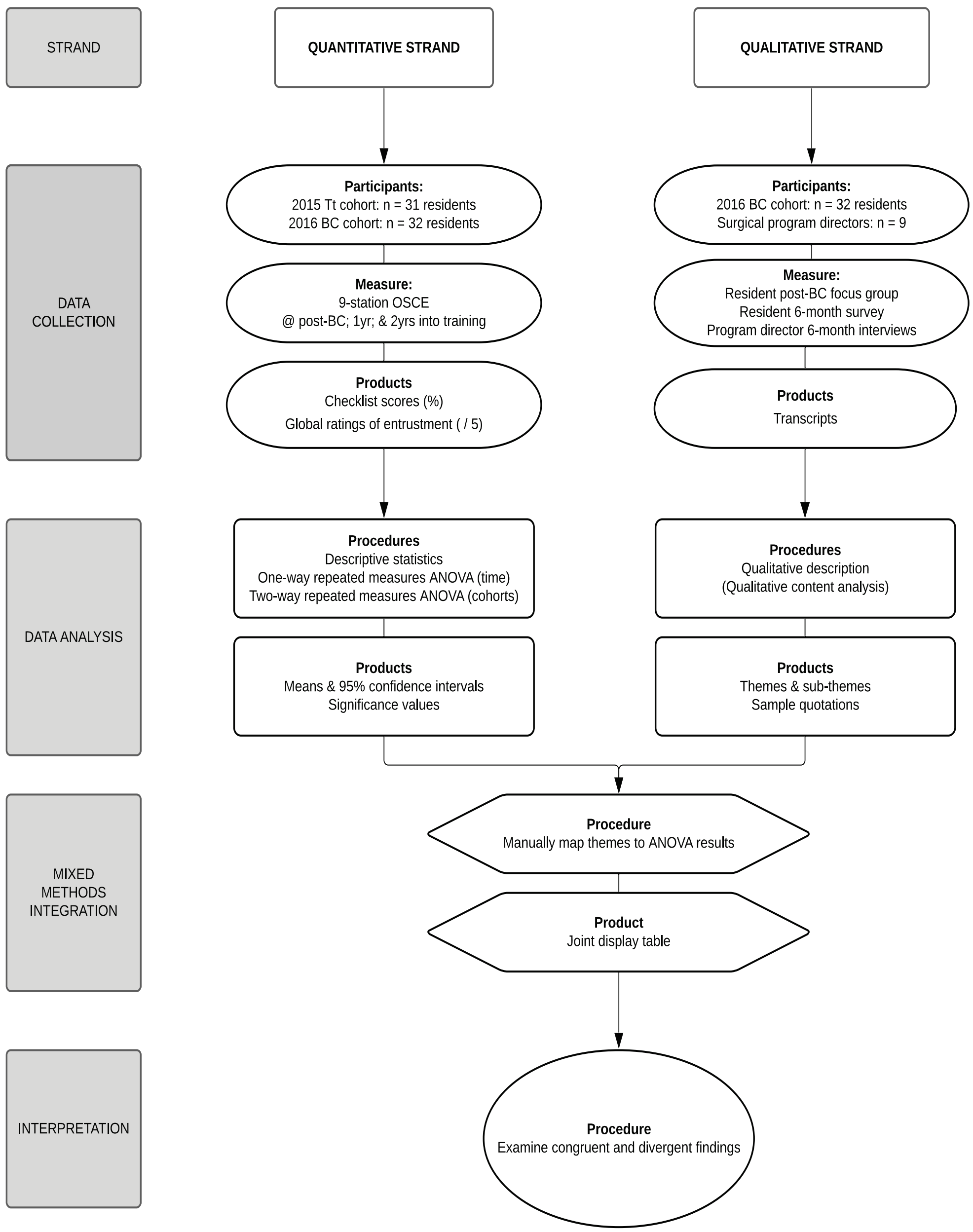


Figure 1. Convergent Parallel Mixed Methods (CPMM) design

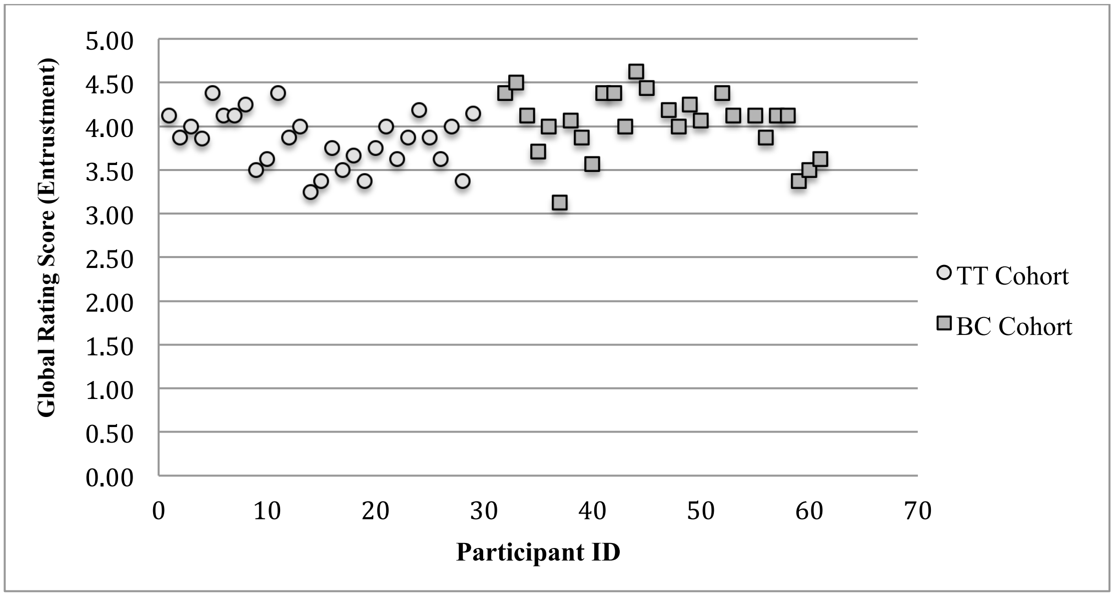


Figure 2. Example Spread of OSCE scores - Entrustment at Two Years into Training

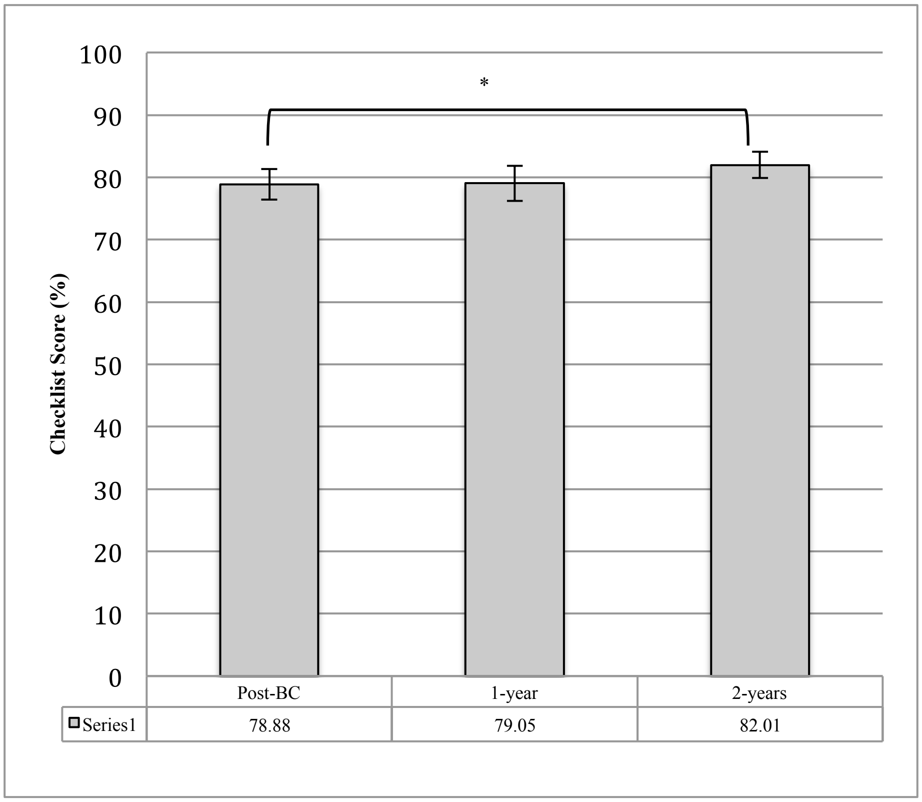
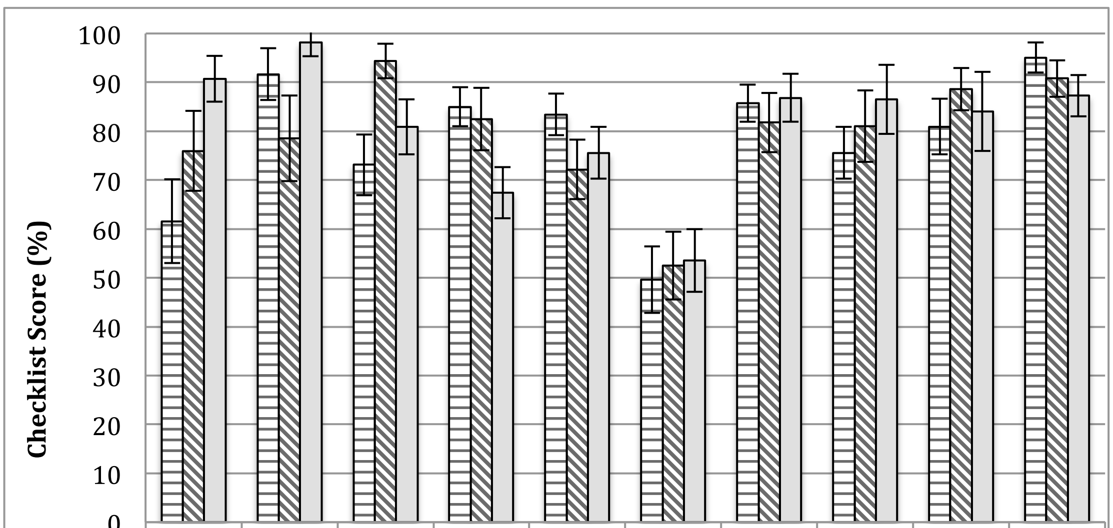


Figure 3. OSCE Checklist Scores for the 2016 Boot Camp Cohort Across Time
* = Statistically significant difference between time points
Error bars show 95% confidence interval



	Informed Consent	Suturing	PreOp Prep	Instruments	Radiation Safety	Electrosurgery	Breaking Bad News	OR Conduct	Recog. Crit. Ill Patient	Handover
Post_BC	61.6	91.67	73.12	85	83.44	49.65	85.76	75.57	80.94	95.09
1year	75.96	78.57	94.34	82.5	72.14	52.5	81.78	81.07	88.57	90.81
2years	90.74	98.15	80.9	67.41	75.56	53.52	86.83	86.53	84.07	87.3

Figure 4. OSCE Checklist Scores (by Station) for the 2016 Boot Camp Cohort Across Time

Error bars show 95% confidence interval

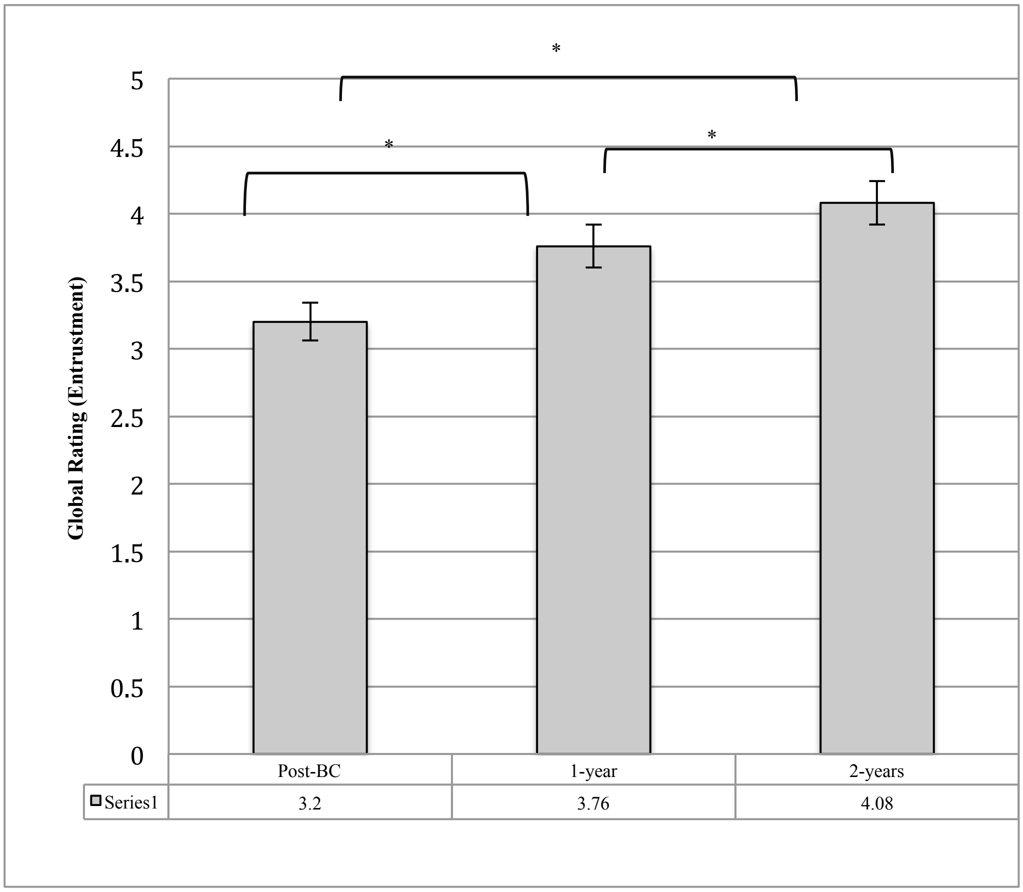
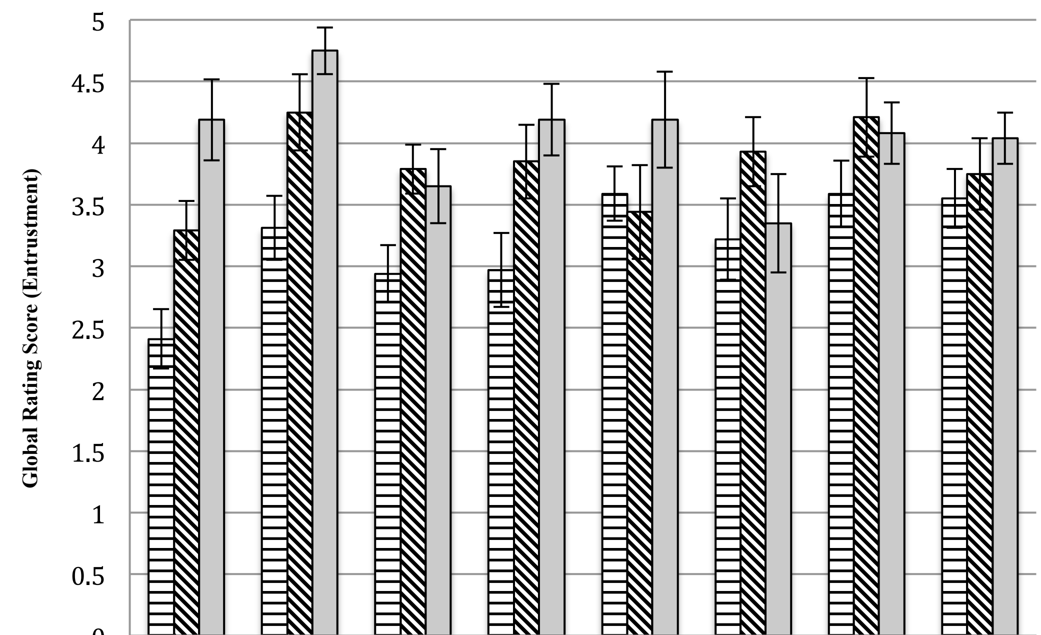


Figure 5. OSCE Entrustment Scores for the 2016 Boot Camp Cohort Across Time
* = Statistically significant difference between time points
Error bars show 95% confidence interval



	Informed Consent	Suturing	PreOp Prep	Breaking Bad News	OR Conduct	Recog. Crit. Ill Patient	Handover	Hand Ties
Post_BC	2.41	3.31	2.94	2.97	3.59	3.22	3.59	3.55
1year	3.29	4.25	3.79	3.85	3.44	3.93	4.21	3.75
2years	4.19	4.75	3.65	4.19	4.19	3.35	4.08	4.04

Figure 6. OSCE Entrustment Scores (by Station) 2016 Boot Camp Across Time
Error bars show 95% confidence interval

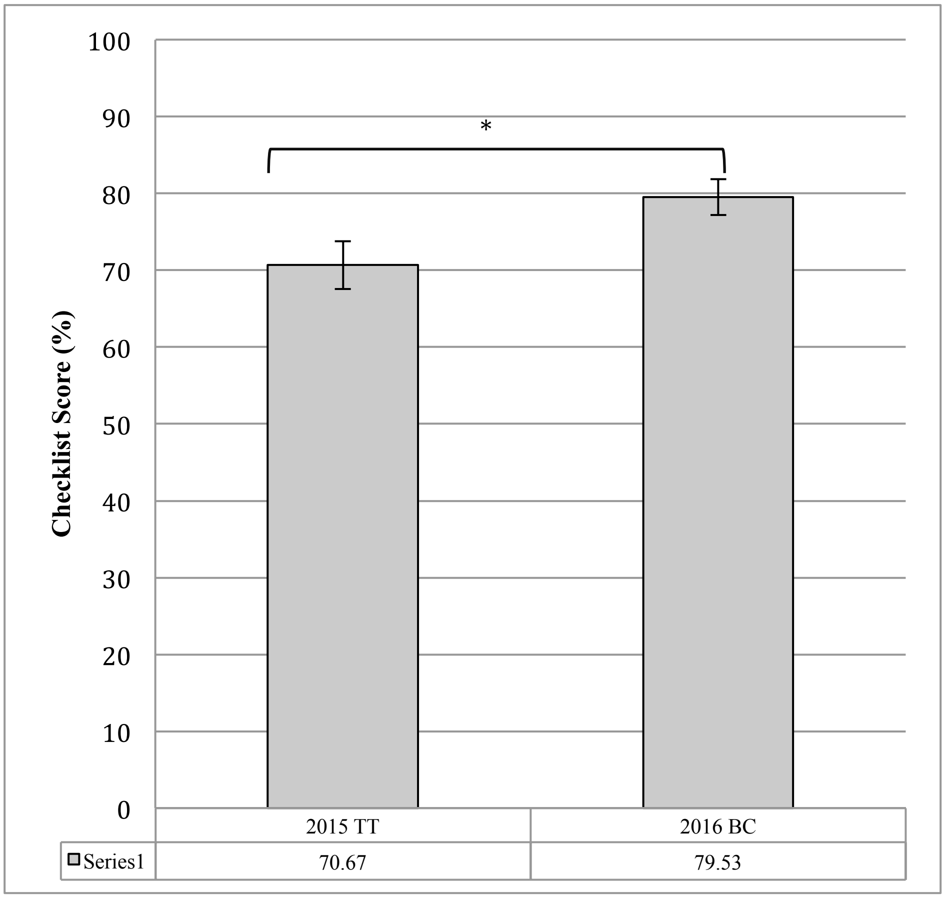
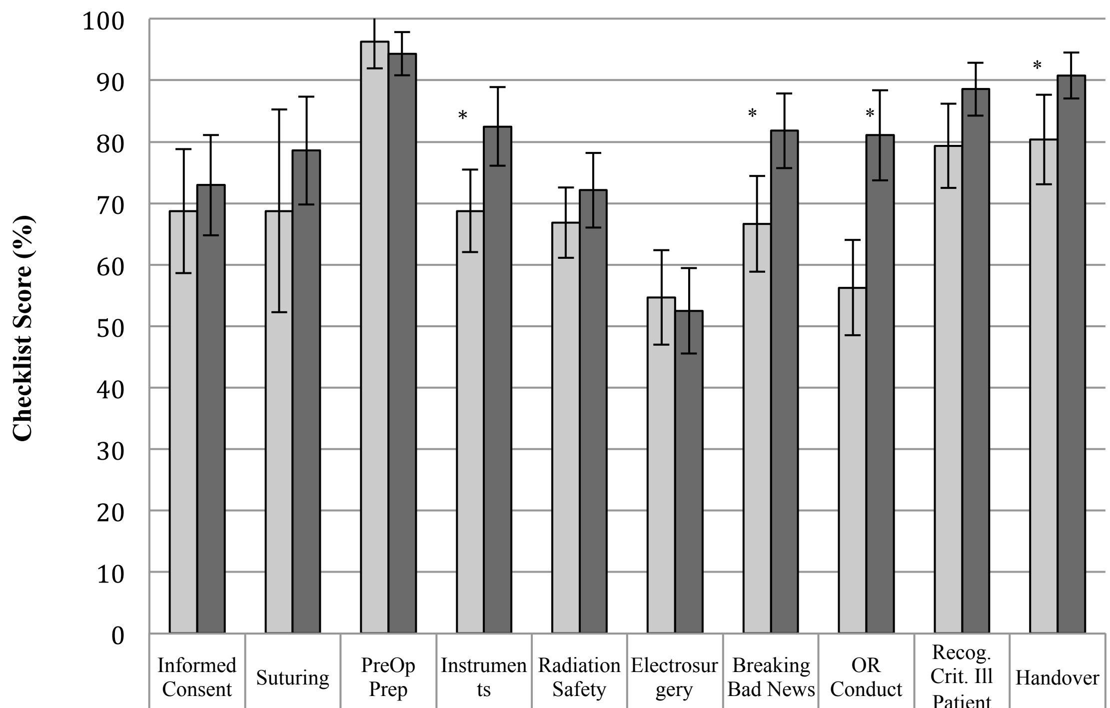


Figure 7. OSCE Checklist Scores for the 2015 Traditionally Trained and 2016 Boot Camp Cohorts at One Year Into Training

* = Statistically significant difference between cohorts
Error bars show 95% confidence interval



	Informed Consent	Suturing	PreOp Prep	Instrumen ts	Radiation Safety	Electrosur gery	Breaking Bad News	OR Conduct	Recog. Crit. Ill Patient	Handover
2015_TT	68.75	68.75	96.25	68.75	66.87	54.69	66.67	56.25	79.37	80.36
2016_BC	72.96	78.57	94.34	82.5	72.14	52.5	81.78	81.07	88.57	90.81

Figure 8. OSCE Checklist Scores (by Station) for the 2015 Traditionally Trained and 2016 Boot Camp Cohorts at One Year Into Training

* = Statistically significant difference between cohorts
Error bars show 95% confidence interval

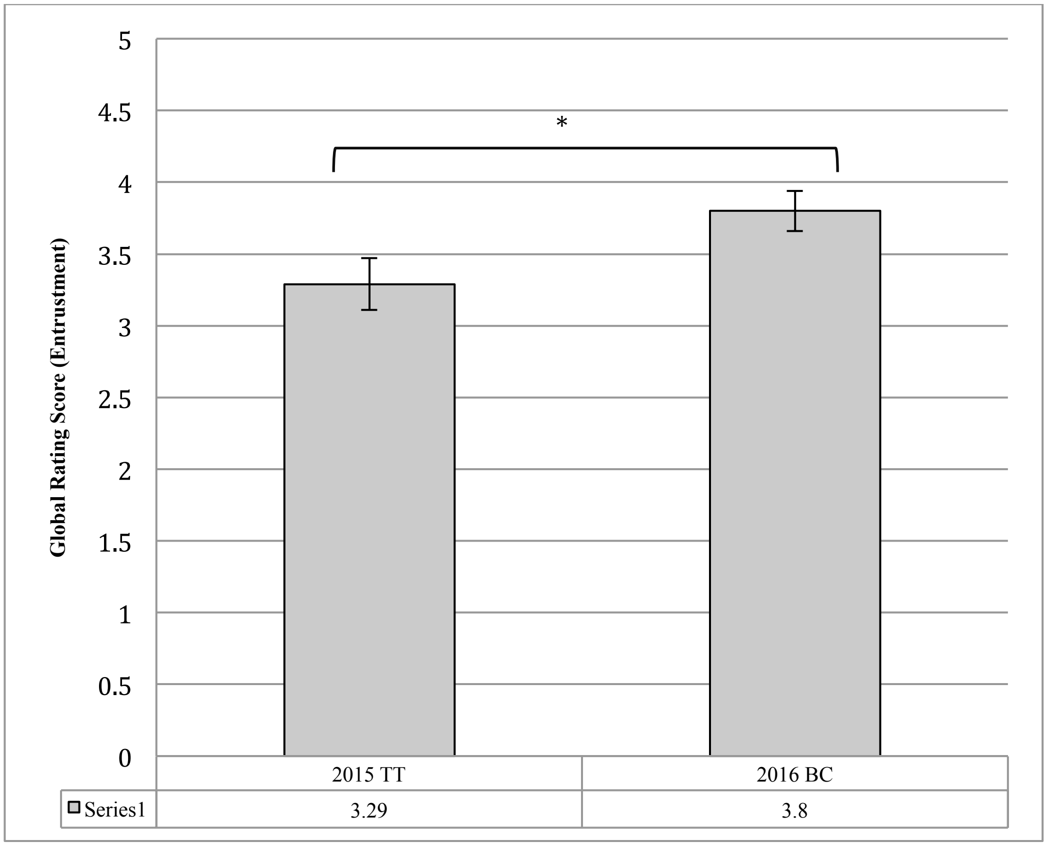
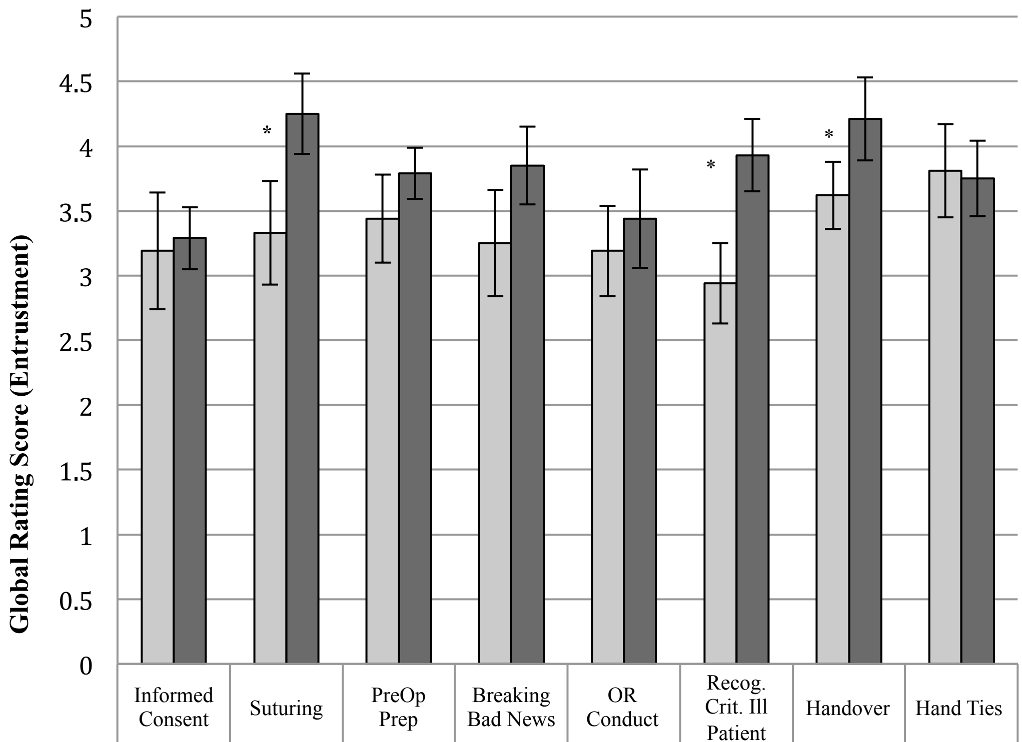


Figure 9. OSCE Entrustment Scores for the 2015 Traditionally Trained Cohort and 2016 Boot Camp Cohorts at One Year Into Training

* = Statistically significant difference between cohorts
Error bars show 95% confidence interval



	Informed Consent	Suturing	PreOp Prep	Breaking Bad News	OR Conduct	Recog. Crit. Ill Patient	Handover	Hand Ties
2015_TT	3.19	3.33	3.44	3.25	3.19	2.94	3.62	3.81
2016_BC	3.29	4.25	3.79	3.85	3.44	3.93	4.21	3.75

Figure 10. OSCE Entrustment Scores (by Station) for the 2015 Traditionally Trained Cohort and 2016 Boot Camp cohort at One Year Into Training

* = Statistically significant difference between cohorts
Error bars show 95% confidence interval

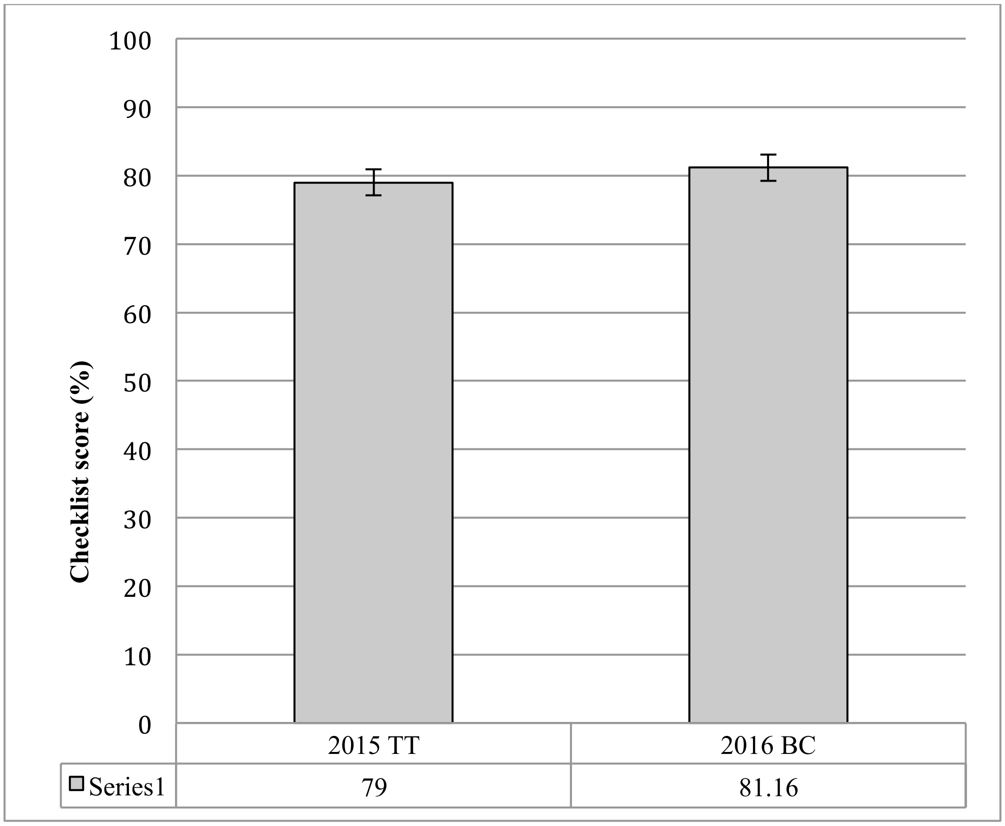
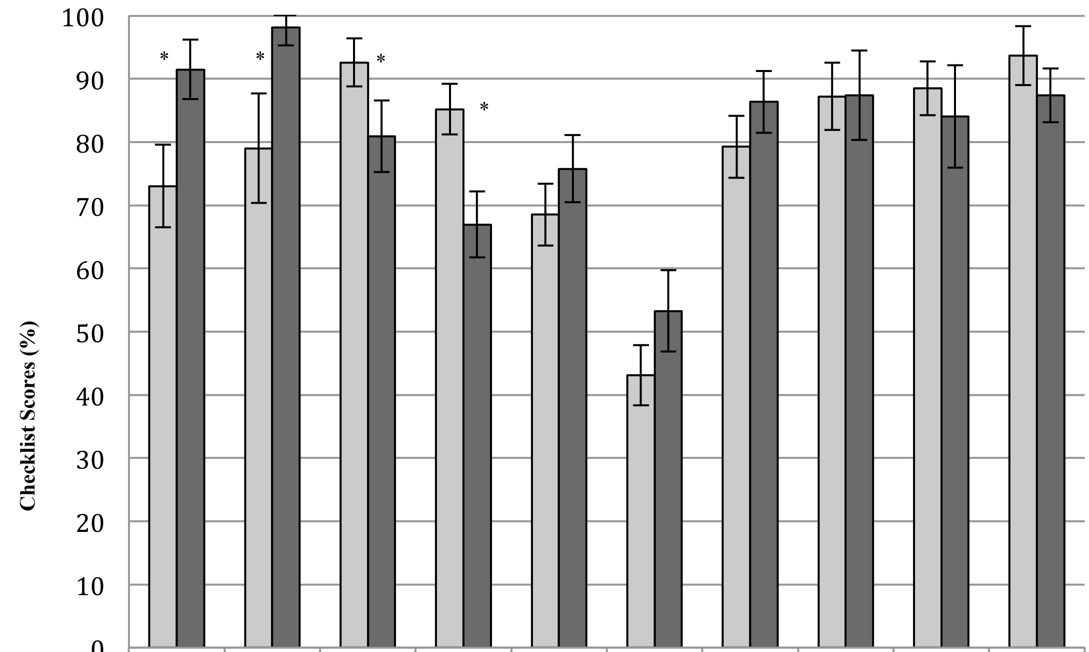


Figure 11. OSCE Checklist Scores for the 2015 Traditionally Trained and 2016 Boot Camp Cohorts at Two Years Into Training

Error bars show 95% confidence interval



	Informed Consent	Suturing	PreOp Prep	Instruments	Radiation Safety	Electrosurgery	Breaking Bad News	OR Conduct	Recog. Crit. Ill Patient	Handover
2015_TT	73.01	79.01	92.59	85.19	68.52	43.11	79.26	87.22	88.52	93.65
2016_BC	91.48	98.08	80.9	66.92	75.77	53.27	86.33	87.41	84.04	87.36

Figure 12. OSCE Checklist Scores (by Station) for the 2015 Traditionally Trained and 2016 Boot Camp Cohorts at Two Years Into Training

* = Statistically significant difference between cohorts

Error bars show 95% confidence interval

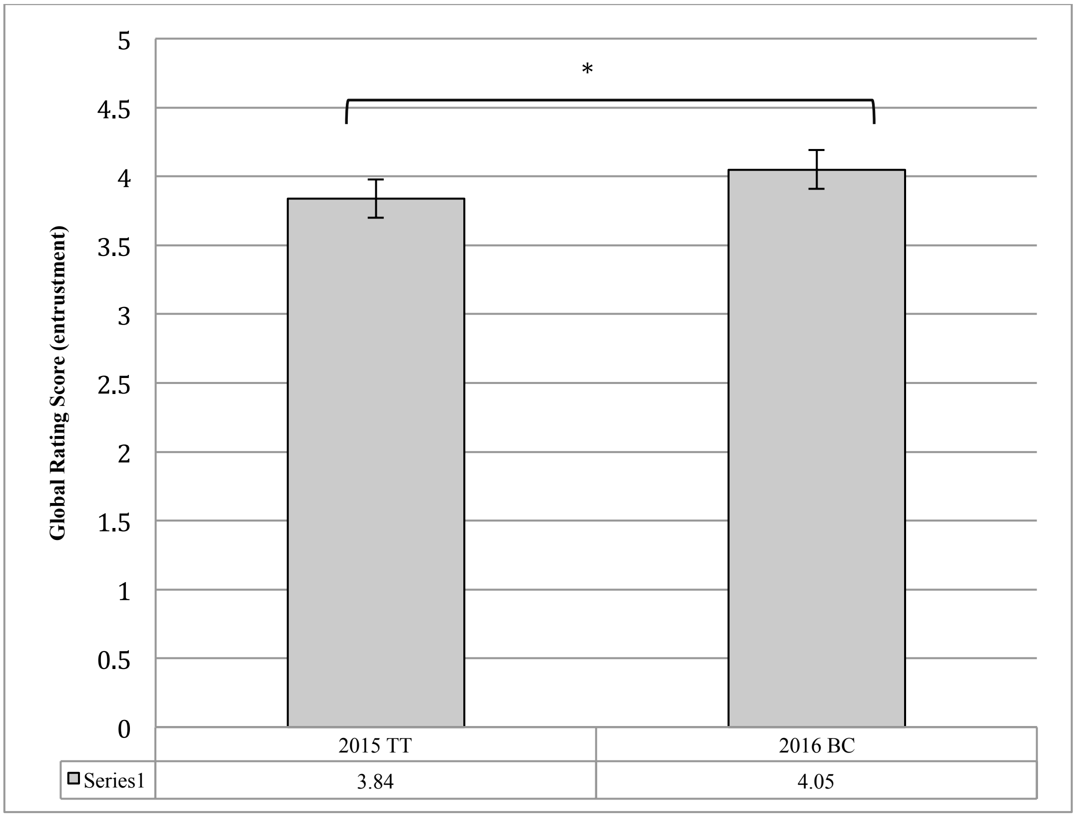
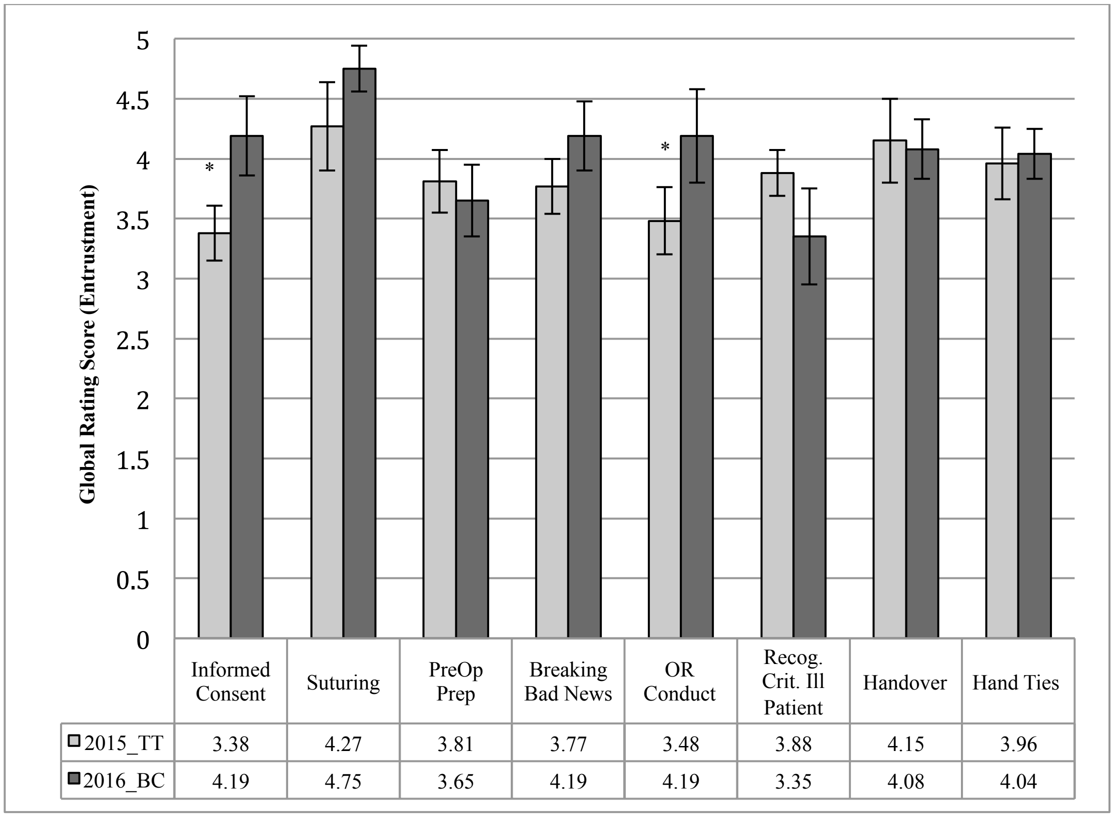


Figure 13. OSCE Entrustment Scores for the 2015 Traditionally Trained and 2016 Boot Camp Cohorts at Two Years Into Training

* = Statistically significant difference between cohorts
Error bars show 95% confidence interval



	Informed Consent	Suturing	PreOp Prep	Breaking Bad News	OR Conduct	Recog. Crit. Ill Patient	Handover	Hand Ties
2015_TT	3.38	4.27	3.81	3.77	3.48	3.88	4.15	3.96
2016_BC	4.19	4.75	3.65	4.19	4.19	3.35	4.08	4.04

Figure 14. OSCE Entrustment Scores (by Station) for the 2015 Traditionally Trained and 2016 Boot Camp Cohorts at Two Years Into Training

* = Statistically significant difference between cohorts

Error bars show 95% confidence interval

Appendix A

Sample OSCE Assessment Form

Station 1: Informed Consent

Instructions to Candidates:

A 65 y/o male 5 days post-op from an open sigmoid resection with primary anastomosis for diverticular disease has developed a wound infection with a 5 x 3 cm area of erythema and fluctuance. There is no discharge from the wound and the skin staples are still in place. Please obtain informed consent from the patient to incise and drain the wound abscess.

ITEM	**Not done, Incorrect**	**Done, Correct**
1. Explained the rationale for the procedure	**0**	**1**
2. Accurately described the procedure	**0**	**1**
3. Described the risks of the procedure	**0**	**1**
4. Described the benefits of the procedure	**0**	**1**
5. Confirmed the patient understood the procedure	**0**	**1**
6. Asked if the patient had any questions	**0**	**1**
7. Correctly completed one of the consent forms provided	**0**	**1**

MAXIMUM TOTAL SCORE **/7**

Please circle the number corresponding to the candidate's performance regardless of the candidate's level of training.

Knowledge of Procedure				
1	**2**	**3**	**4**	**5**
Deficient knowledge; required assistance to explain steps of the procedure		Knew most steps of the procedure		Explained all steps of the procedure without prompting
Communication				
1	**2**	**3**	**4**	**5**
Difficulties verbalizing information; lacked clarity		Adequately verbalized information		Clearly and accurately verbalized information
Professionalism				
1	**2**	**3**	**4**	**5**
Did not listen to the patient's concerns		Listened to most of the patient's concerns		Listened to all of the patient's concerns
Entrustment				
1	**2**	**3**	**4**	**5**
Requires complete supervision/guidance		Requires some supervision/prompting		Requires no supervision, ready for independent practice

Appendix B

Resident Post-Boot Camp Focus Group Interview Guide

Pre-amble

The purpose of this focus group is to gather information on how effective the boot camp was and how it can be *improved f*or the future. No names will be recorded, so every comment will be reported anonymously. The graduate student researcher (N.W.) will listen to the recording and transcribe verbatim. General themes across all 4 focus groups will be reported back to the Department of Surgery leadership to make changes to subsequent boot camps. This is the chance to be as honest as possible; your feedback heavily influences what the first few weeks will look like for next year's cohort. We have based these questions off of your feedback during the OSCE.

Prompts:

1. Do you think there is a gap between medical school and residency? If so, why?
2. Would you change the length of boot camp? Shorter? Longer?
3. Would you change the timing of boot camp (i.e. not at the beginning of July)?
4. Did you like the format (didactic morning, technical skills afternoon)? What would you change?
5. Did you use the source documents? If so when did you review them? Before the session or just before the OSCE?
6. How much studying did you do outside of boot camp hours? Was it for surgical foundations or specialty specific? If specialty specific, was this required by your programs?
7. Do you think the "SPs" would have been more useful had they had formal training? If so, why?
8. Peer feedback was reported as less helpful than resident/staff feedback. Why? How could that be improved?
9. Do you think watching others perform the technical skills was valuable? Why or why not?

10. 50% of the group reported the breaking bad news session was not helpful, however the OSCE evaluators thought there is room for major improvement on that skill. How could we make that session more helpful?

11. Please refer to the list of other sessions & give feedback for how each one could be improved.

12. We received a lot of feedback saying the boot camp would have been much more useful if it was specialty specific. Why do you think this is? Is there a place for both general and specialty specific skills?

13. How would you incorporate specialty specific components into the current boot camp?

14. What else would you add to the curriculum?

15. Do you think the OSCE was fair? What would you change?

16. Any other suggestions for how we could evaluate effectiveness?

17. Overall did the boot camp decrease or increase anxiety about starting residency?

18. Overall was the boot camp useful in preparing you for residency?
    a. Specifically was the boot camp useful for everyone, or just those joining McMaster for the first time?
    b. How could we make it more useful for Mac grads while still addressing necessary administrative stuff?

19. Would you recommend running a boot camp again next year?

20. Any additional comments or suggestions?

END OF SCRIPT

Appendix C

Six Month Follow Up Boot Camp Participant Survey

Dear PGY1s,

Firstly, we would like to thank you for all the feedback you provided during the Surgical Foundations boot camp. We truly value this information and have taken all comments into consideration while planning the 2017 boot camp.

Now that you are 6-months into training, we wanted to conduct a quick follow up survey. Please answer all questions. The information provided to the research team will remain completely anonymous.

1. Do you believe that having a two week boot camp in July made switching into the role of a resident easier? Please explain why or why not.
2. We have attached the schedules from the 2016 boot camp, and also the preliminary draft of the 2017 schedule. Please review the 2017 schedule and provide feedback on the changes we have made.
3. Are there any additional topics/skills you have encountered over the past 6 months that you think would be beneficial to include in the 2017 boot camp?
4. Did having the opportunity to work with some staff and senior residents before starting clinical duties make the transition from medical school to residency easier?
5. Do you believe the two week boot camp is a valuable activity and should be continued with future cohorts? Please explain why or why not.

Thank you for your participation

If you have any questions or wish to become involved in the 2017 boot camp, please contact Natalie Wagner (wagnernk@mcmaster.ca)

END OF SURVEY

Appendix D

Program Director Interview Guide

1. What is your specific area of practice?
    a. For how long have you been practicing?
2. Did you participate in the Surgical Foundations boot camp? If so, what session(s)? If not, are there are particular reasons that you did not participate, that you would care to share?
3. What was your general impression surrounding the usefulness of the boot camp?
4. Have you heard any feedback (either positive or negative) from staff and/or residents on the usefulness of the boot camp? Please explain.
5. Have you worked with a PGY1 in the last 6 months? If so, how frequently?
6. Have you noticed any positive effects of the Surgical Foundations boot camp? Please explain
7. Have you noticed any negative effects of the Surgical Foundations boot camp? Please explain.
8. Specifically, do you believe the Surgical Foundations boot camp helped prepare the PGY1s for clinical duties? If so, please explain.
9. Have you noticed a difference in teamwork between this PGY1 cohort and previous years?
10. Do you believe the two week boot camp is a valuable activity and should be continued with future cohorts? Please explain why or why not.
11. Do you have any final comments regarding the boot camp?
12. Lastly, are there any specific skills you believe should be included in the Foundations program, either in the boot camp or throughout the year?

Thank you for participating.

END OF SCRIPT

4.7 References

Blackmore, C., Austin, J., Lopushinsky, S. R., & Donnon, T. (2014). Effects of Postgraduate Medical Education "Boot Camps" on Clinical Skills, Knowledge, and Confidence: A Meta-Analysis. *J Grad Med Educ, 6*(4), 643-652. doi:10.4300/JGME-D-13-00373.1

Blanca, M. J., Alarcon, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552-557. doi:10.7334/psicothema2016.383

Bland, C., Starnaman, S., Wersal, L., Moorhead-Rosenberg, L., Zonia, S., & Henry, R. (2000). Curricular Change in Medical Schools: How to Suceed. *Acad Med, 75*, 575 - 594. Retrieved from http://ovidsp.dc1.ovid.com.libaccess.lib.mcmaster.ca/.

Brydges, R., Hatala, R., Zendejas, B., Erwin, P. J., & Cook, D. A. (2015). Linking simulation-based educational assessments and patient-related outcomes: a systematic review and meta-analysis. *Acad Med, 90*(2), 246-256. doi:10.1097/ACM.0000000000000549

Cohen, D., & Crabtree, B. (2006). Qualitative research guidelines project. Retrieved from http://www.qualres.org/HomeSemi-3629.html

Creswell, J. W. (2007). *Qualitative Research Inquiry Methods; Second Edition*. Thousand Oaks, CA: SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and Conducting Mixed Methods Research* (3 ed.). Thousand Oaks, CA: SAGE Publications.

Cypress, B. S. (2017). Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations. *Dimens Crit Care Nurs, 36*(4), 253-263. doi:10.1097/DCC.0000000000000253

Grierson, L. E. (2014). Information processing, specificity of practice, and the transfer of learning: considerations for reconsidering fidelity. *Adv Health Sci Educ Theory Pract, 19*(2), 281-289. doi:10.1007/s10459-014-9504-x

Harden, R. M., Stevenson, M., Wilson Downie, W., & Wilson, G. M. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal, 1*, 447-451. Retrieved from https://www-jstor-org/.

Hauer, K. E., Ten Cate, O., Boscardin, C., Irby, D. M., Iobst, W., & O'Sullivan, P. S. (2014). Understanding trust as an essential element of trainee supervision and learning in the workplace. *Adv Health Sci Educ Theory Pract, 19*(3), 435-456. doi:10.1007/s10459-013-9474-4

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qual Health Res, 15*(9), 1277-1288. doi:10.1177/1049732305276687

Khan, M. E., Anker, M., Patel, B. C., Barge, S., Sadhwani, H., & Kohle, R. (1991). The use of focus groups in social and behavioural research: some methodological issues. *World Health Stat Q, 144*(3), 145-149. Retrieved from https://www.researchgate.net/.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. . *Personnel Psychology, S8*(4), 981-1007. Retrieved from http://ink.library.smu.edu.sg/lkcsb_research/5679.

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*, 273-278. doi:10.1002/bjs.1800840237

Neergaard, M. A., Olesen, F., Andersen, R. S., & Sondergaard, J. (2009). Qualitative description - the poor cousin of health research? *BMC Med Res Methodol, 9*, 52. doi:10.1186/1471-2288-9-52

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Med Educ, 46*(7), 636-647. doi:10.1111/j.1365-2923.2012.04243.x

Pugh, D., Touchie, C., Humphrey-Murto, S., & Wood, T. J. (2016). The OSCE progress test--Measuring clinical skill development over residency training. *Med Teach, 38*(2), 168-173. doi:10.3109/0142159X.2015.1029895

Rekman, J., Gofton, W., Dudek, N., Gofton, T., & Hamstra, S. J. (2016). Entrustability Scales: Outlining Their Usefulness for Competency-Based Clinical Assessment. *Acad Med, 91*(2), 186-190. doi:10.1097/ACM.0000000000001045

Sandelowski, M. (2000). Whatever happened to Qualitative Description? *Research in Nursing & Health, 23*, 334-340. Retrieved from https://journals-scholarsportal-info.libaccess.lib.mcmaster.ca/.

Sonnadara, R. R., Garbedian, S., Safir, O., Nousiainen, M., Alman, B., Ferguson, P., . . . Reznick, R. (2012). Orthopaedic Boot Camp II: examining the retention rates of an intensive surgical skills course. *Surgery, 151*(6), 803-807. doi:10.1016/j.surg.2012.03.017

van der Leeuw, R. M., Lombarts, K., Onyebuchi, A. A., & Heinerman, M. J. (2012). A systematic review of the effects of residency training on patient outcomes. *BMC Medicine, 10*(65), 1-11. Retrieved from http://www.biomedcentral.com/1741-7015/10/65.

Chapter 5

The Objective Structured Clinical Examination as an Indicator of Clinical Performance

5.0 The Objective Structured Clinical Examination as an Indicator of Clinical Performance

5.1 Introduction

As discussed in chapter four, the Objective Structured Clinical Examination (OSCE) has been considered the gold standard of assessment in medical training for the past several decades (Pugh, Touchie, Humphrey-Murto, & Wood, 2016). However, the results from our study on the long-term implications of a boot camp (BC) program suggest that OSCE scores might not always be an accurate index of clinical performance (chapter 4.4.2). To explore this idea further, chapter five describes a study comparing OSCE assessments with assessments completed in the clinical environment. In order to understand the context in which this study occurs, we begin with a brief overview of the history of OSCEs, discuss existing research on the psychometric properties of OSCEs, explore the challenges with measuring performance in the clinical environment, and suggest new opportunities to assess clinical performance with the introduction of competency-based medical education (CBME).

5.1.1 Evolution of the Objective Structured Clinical Examination (OSCE) in Canada

In the mid-1950s, oral and written examinations were popular assessment methods in medical training (Cowles, 1954). However, in the 1960s, educators started to express concerns that oral examinations were too subjective, written examinations only provided information on trainees' medical knowledge, and that neither examination format truly assessed the clinical skills or competencies that are required of a physician (Charvat, McGuire, & Parsons, 1968; McGuire, 1966). In response, educators at University of Dundee, Scotland created the Objective Structured Clinical Examination (OSCE), a bell-ringer type examination in which candidates rotate through various clinical scenarios (stations). The original OSCE consisted of eight pairs of stations, each station five-minutes in length (Harden, Stevenson, Wilson Downie, & Wilson, 1975). In the first station, medical trainees were given written instructions and asked to complete a task or procedure, for example, taking a history or physical examination from a standardized patient. In the second station of the pair, medical trainees had to answer questions about their findings and interpretations from the preceding station. Trainees were scored on each station using a checklist, and their final OSCE score was a combination of checklist scores from all stations (Harden et al., 1975). As the OSCE addressed many of the concerns with oral and written examinations, medical educators around the world were quick to adopt this new method (Hodges, 2007).

In 1979, Ian Hart, a Canadian physician who had spent his sabbatical with Harden during the 1980s, introduced the OSCE at the University of Ottawa (Hart, 1980). As the chief promoter of OSCEs in Canada, Ian Hart established a number of research initiatives and conferences to encourage scholarship of this examination format. As a result of these efforts, Canada became one of the leading countries in OSCE research (Grand'Maison, Lescop, & Brailovsky, 1993). In 1989 the Medical Council of Canada (MCC), the licensing body for all physicians in Canada, embarked on a pilot project to include OSCEs as part of their licensing examinations (Reznick et al., 1992). The MCC pilot OSCE consisted of 20 stations and took place across four sites, involving a total of 240 medical graduates. Results from this pilot study suggested that OSCE scores had a low correlation with written examination scores (r = .32 - .35); however, the overall reliability was moderate (r = .54 - .60), mean OSCE scores did not vary by site (Toronto or Montreal), or delivery language (English or French), and examiners felt the OSCE stations were important, challenging, and suitable for testing candidates' core knowledge (Reznick et al., 1992). Based on these results, the authors concluded that OSCEs could be an extremely valuable adjunct to written examinations (Reznick et al., 1992). In 1995, the OSCE became part of the official licensing requirements for all physicians in Canada (Dauphinee, 1995).

5.1.2 Creation of the Objective Structured Assessment of Technical Skills

While the inclusion of the OSCE as a licensing requirement in Canada helped ensure that candidates had attained adequate competency in non-technical domains such as clinical knowledge, communication, and professionalism before they completed their training, it did little to ensure that trainees had developed sufficient technical skills. To address this critical gap, Martin et al. (1997) developed the Objective Structured Assessment of Technical Skills (OSATS). The OSATS format was very similar to that of an OSCE, with learners rotating through multiple simulated stations; however, the OSATS introduced a new scoring system. While the OSCE used a task-specific checklist for each station, the OSATS used both task-specific checklists for each station, as well as seven, five-point global rating scales that were common across all stations. The results from Martin et al.'s study presented compelling data that using both checklists and global rating scales resulted in moderate inter-rater reliability (Intraclass correlation coefficient

(ICC)= .64 – .72) and moderate inter-station reliability, or the relationship between stations (α = .61 – .74) (Koo & Li, 2016). Moreover, their study suggested that these global rating scales were able to discriminate between trainees who were at different stages of training. Martin et al.'s new scoring system was quickly adopted for use in a wide variety of assessments in medical training, including the OSCE.

5.1.3 Psychometric Properties of Objective Structured Clinical Examinations

Since the introduction of the OSATS scoring system, there have been a number of publications on the psychometric properties of OSCEs. While global rating scales have been found to be superior to task-specific checklists in general (Regehr, Macrae, Reznick, & Szalay, 1996), studies suggest that both checklist and global rating scales demonstrate moderate to high inter-rater reliability, and are able to discriminate between level of training when used to measure performance on the OSCE (Hodges, Turnbull, Cohen, Bienenstock, & Norman, 1996; Pugh et al., 2015).

A recent review by Hatala, Cook, Brydges, and Hawkins (2015) suggested that using the OSATS, specifically the global rating scale components, to measure technical skills in simulation demonstrates high inter-rater reliability (ICC > .70), a strong relationship with level of expertise (novice versus expert), a strong relationship with other performance measures in simulation, and is able to capture improvement with training when assessing technical skills in simulation (Hatala et al., 2015). However, the authors also identified a lack of information regarding the relationship between OSATS scores and clinical performance. As such, the authors conclude that there is strong evidence for using the OSATS as a formative assessment, but further work is needed to use the OSATS for scoring for summative assessments.

We recently conducted a systematic review on studies comparing OSCE scores (on both technical and non-technical skills) to clinical performance (Wagner, Fahim, & Sonnadara, In preparation). Of the 2266 identified articles, 40 were included in the review. Of those papers, 45% found a significant correlation between clinical performance and OSCE scores; 17.5% found no significant correlations, and 30% had mixed results (Wagner et al., In preparation). As the OSCE and clinical assessments often focused on different skills, used different scales, had different raters, and allowed for variability in the amount of time between assessments, we proposed that conflicting findings might be due to heterogeneity within and between studies. In order to understand the relationship between OSCE scores and clinical performance, further research is needed.

5.1.4 Challenges Assessing Clinical Performance

Unfortunately, trainees' clinical performance is quite difficult to measure. Typically, studies have used Workplace-based Assessments (WBAs), or observations of trainees in 'real-life' professional settings, to capture clinical performance (Govaerts, Van de Wiel, Schuwirth, Van der Vleuten, & Muijtjens, 2013). Upon their inception, many different types of WBAs were developed, including: the mini-clinical evaluation exercise (mini-CEX), clinical encounter cards, Direct Observation of Procedural Skills (DOPS), and multi-source feedback (MSF) (Norcini & Burch, 2007). While these WBAs have great potential in documenting trainee competence and have been implemented in the

United Kingdom as part of the Intercollegiate Surgical Curriculum Programme (Beard & Bussey, 2007), their adoption has faced a number of challenges.

Workplace-based Assessments have been criticized for their low inter- and intra-rater reliability (Govaerts et al., 2013). Other studies have also found that faculty supervisors are not willing to assign low marks to poorly performing trainees due to insufficient observation, fear of appeals and legal action, and fear of reciprocated poor faculty evaluations (McQueen et al., 2016). Dudek, Marks, and Regehr (2005) proposed that the perception that WBAs were being used for summative decisions has created a 'failure to fail' phenomenon. As a result of this 'failure to fail' phenomenon, some studies have suggested that educators perceive WBAs as a 'tick-box' exercise, rather than a useful assessment (e.g. Massie & Ali, 2016). In addition to conflicting evidence regarding their effectiveness, there are growing pressures on physicians to be as efficient as possible in healthcare delivery (chapter 1.4). This means that WBAs are often not being completed in the first place. Thus, it is not surprising that there is a lack of evidence on whether clinical performance (typically measured by WBAs) is correlated with OSCE performance.

5.1.5 Workplace-based Assessments and Competency-Based Medical Education

Fortunately, the global shift towards competency-based medical education (CBME) frameworks has renewed interest in refining WBAs. In CBME, trainees' progress is based on the demonstration of competence, rather than time spent in training. As a result, many CBME models, including Competence By Design (CBD), the Canadian model being implemented by the Royal College of Physicians and Surgeons of Canada

(RCPSC), mandate regular assessment of 'in situ' trainee competence. In the Canadian CBD model, the RCPSC has determined that no single assessment should impact a trainees' progression through residency. Rather, regular assessments should act as an opportunity to provide trainees with feedback in a timely manner. When all of the assessments are brought together in a portfolio, data points can be aggregated to provide educators with a sense of the trainees' overall competence (van der Vleuten et al., 2012). The hope is that this assessment format changes perceptions on the summative nature of assessments in the clinical environment and mitigates some of the 'failure to fail' phenomenon that has previously influenced WBAs. In addition to this change in how assessments are being used, there has been a movement towards using 'entrustment' anchors, rather than 'performance-based' anchors to capture competence on WBAs (Rekman, Gofton, Dudek, Gofton, & Hamstra, 2016). For example, the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE), which has been adopted by many Canadian training programs, includes a global rating scale with anchors such as "I had to do" or "I didn't need to be there", rather than "poor" or "excellent". Recent literature suggests this type of entrustment scale is more transparent, facilitates formative feedback, and promotes the use of the entire scale (Dudek, Gofton, Rekman, & McDougall, 2019). Entrustment anchors have also been found to be more intuitive, which makes it easier for supervisors to complete these WBAs (Dudek et al., 2019).

In turn, these changes to assessment have the potential to improve WBAs. Educators are now incentivized to complete WBAs. The emphasis on frequent formative assessment might mitigate the failure to fail phenomenon. The advent of entrustment-based WBAs makes the process of capturing performance less burdensome for medical educators, while at the same time improving the accuracy of data. As a result of these changes, we now have a unique opportunity to build on previous research and explore the relationship between OSCE performance and clinical performance in ways that have previously not been possible.

5.1.6 Study Objectives

The purpose of this study was to compare novice surgical trainee performance measured through an OSCE at the onset of residency with in situ performance measured through WBAs within the first month of clinical training. As part of this work, we also explored how these measures related to other formative assessments completed in simulation.

5.2 Methods

5.2.1 Setting and Participants

This study took place at McMaster University in Hamilton, ON. Participants were first-year residents enrolled in the McMaster Surgical Foundations (SF) program in 2017 (n = 32). As discussed in chapter three and four, the McMaster SF program is comprised of first- and second-year surgical residents from nine different specialties, including: Cardiac Surgery, Plastic Surgery, Obstetrics and Gynecology, Vascular Surgery, Urology, Otolaryngology-Head and Neck Surgery, Neurosurgery, General Surgery, and Orthopaedics (Table 1). In 2017, all first-year residents in the SF program completed the second iteration of the McMaster SF boot camp (BC), at the onset of residency (chapter 3.7). Following the two-week BC, residents completed clinical rotations at the direction of their home specialty. Clinical rotations took place across four different hospital sites in Hamilton, ON. Starting in August 2017, residents met for SF teaching sessions (academic half-days), once per week. During the academic half-days, residents participated in didactic, case-based, or lab-based/practical skills sessions and were free from their clinical responsibilities.

5.2.2 Study Design

This study took place from July 3, 2017 – August 6, 2017. Data collection occurred in three stages (Table 2). All assessments collected in this study were designed by the research team and incorporated as part of the SF training program at McMaster. In order to be able to use these data for research purposes, the research team obtained consent from the SF program to access and de-identify the assessment data.

Surgical Foundations 2017 boot camp.

As part of the BC, residents (n = 32) participated in three suturing sessions, which took place in the Anatomy Lab at McMaster University. Residents were provided with session objectives, a variety of suturing materials, and pigskin models. Residents were instructed to work through the objectives at their own pace, while volunteer facilitators circulated and provided guidance. Each suturing session included five or six volunteer facilitators from the Department of Surgery. Volunteer facilitators were staff surgeons, fellows, or senior residents. By the end of each session, residents were required to have one facilitator complete an assessment on their suturing.

Suturing assessments were based off the O-SCORE and included several global rating scales and a space for narrative feedback (Appendix A). The research team collected completed suturing assessments at the end of each BC session. Residents were assigned a participant number; all assessment data were de-identified and saved to a secure database.

Post-boot camp OSCE.

On the second last day of the BC, residents participating in the BC completed an OSCE (n = 32). The OSCE was comprised of eight, skill-based stations and one rest station. The eight skill-based stations included: informed consent, pre-operative preparation, suturing, breaking bad news, Foley catheter insertion, recognition of the critically ill patient, handover and hand ties, as well as a multiple-choice quiz on surgical instrumentation, electrosurgery, and radiation safety. The OSCE stations were based on BC content, which was linked to the RCPSC Entrustable Professional Activities (EPAs) for new surgical residents (chapter 1.7.4).

All OSCE scenarios were created by senior residents or faculty from the Department of Surgery and reviewed by the SF Program Director. Volunteer examiners (faculty, fellows, and senior residents from the Department of Surgery) evaluated resident performance on each station using an OSATS-style scoring system, including a task-specific checklist for each station and one global-rating of entrustment common to all stations (Appendix B).

Workplace-based Assessments.

Following the BC, residents started clinical rotations at the direction of their home specialty. During the first four weeks of these clinical rotations, residents were required to complete six workplace-based assessments on at least two of the following skills:

informed consent, breaking bad news, suturing, Foley catheter insertion, or hand ties. These skills were selected as they were topics included in the BC, assessed during the OSCE, and included both technical and non-technical skills. Moreover, the SF Program Director felt residents would be likely to encounter these tasks during the first few weeks of their clinical rotations. The flexibility in assessment completion (six assessments on at least two different skills) was provided due to the unpredictable nature of clinical encounters.

Residents were responsible for initiating the assessments and returning the completed forms to the SF Program Coordinator. Staff surgeons, fellows or senior residents could complete the assessments. To ensure the assessments generated high-quality data, while also being feasible to complete within the clinical environment, the assessment tool was based on the O-SCORE and consisted of one global rating of entrustment and a space for narrative feedback (Appendix C).

5.2.3 Statistical Analyses

To compare residents' performance between the assessment strategies, two sets of analyses were conducted. All analyses were conducted in IBM SPSS (Version 25.0.0.1). Our primary analyses compared OSCE scores with WBA scores. As residents were not required to complete a WBA on all five skills, we were unable to run a repeated-measures ANOVA. Instead, scores for each skill were compared using paired sample t-tests and Pearson correlation coefficients. When residents completed multiple WBAs on a skill, averages were used. When residents completed one WBA on a skill, their individual score was used. When residents did not complete a WBA on a certain skill, they were excluded from the analysis. If assumptions were violated, Wilcoxon signed rank tests and Spearman's rank correlations were used. To account for multiple comparisons, a Bonferroni correction was used. As such, statistical significance was considered when $p < .01$. To confirm the results, the correlation data were also bootstrapped (N = 1000 samples).

As suturing was the only task evaluated during the BC, the OSCE, and the clinical environment, our secondary analyses compared BC formative assessment scores, OSCE scores, and WBA scores using a one-way, repeated-measures ANOVA. Scores from the three BC suturing sessions were averaged for each resident. Workplace-based Assessment scores on suturing were also averaged for residents, where applicable. Those that did not complete WBAs on suturing were excluded from the analysis. A Bonferroni correction was used to account for multiple comparisons. Statistical significance was considered at $p < .05$. In addition to the one-way, repeated measured ANOVA, correlations between the BC formative assessment scores, OSCE scores, and WBA scores for suturing were explored using Pearson's correlation coefficient. If assumptions were violated, a Spearman's rank correlation coefficient was used instead. To account for multiple comparisons, a Bonferroni correction was used. As such, statistical significance was considered when $p < .02$.

5.3 Results

5.3.1 OSCE Scores and Workplace-Based Assessment Scores: All Skills

All five OSCE stations scores violated the normality assumption. As t-tests are impacted by normality violations in small sample sizes (Bridge & Sawilowsky, 1999), nonparametric tests were used in our primary analyses. Table 3 presents a summary of descriptive statistics.

Based on the results from the Wilcoxon signed rank tests we found a statistically significant difference between the OSCE scores and the WBAs for informed consent ($Z = 4.05$, $p < .01$), breaking bad news ($Z = 2.99$, $p < .01$), and Foley catheter insertion ($Z = 2.55$, $p = .01$). We did not find a statistically significant difference for hand ties ($Z = 42.03$, $p = .04$) and suturing ($Z = 1.42$, $p = .16$) (Table 4; Figures 1-5). Based on the results of the Spearman's rank correlation tests we did not find any significant associations between OSCE scores and WBA scores, across all skills. The bootstrapped confidence intervals also supported this finding (Table 4).

5.3.2 Comparing Suturing Scores Across Three Assessment Strategies

The OSCE suturing station violated the normality assumption ($p < .01$). However, we elected to use the one-way repeated-measures ANOVA to compare residents' scores across three assessment strategies (BC formative assessments [average], post-BC OSCE, WBAs [average]) as the histograms appeared fairly normal upon visual inspection and ANOVAs are robust to normality violations (Blanca, Alarcon, Arnau, Bono, & Bendayan, 2017). Our results suggest the sphericity assumption was met ($p = .44$). Based on the ANOVA results we found a statistically significant main effect of assessment strategy ($F (2, 56) = 15.77$; $p < .01$). Through pairwise comparisons we found a statistically significant difference between the formative BC assessment scores and WBA scores [$t (28) = 3.97$, $p < .01$]. We did not find statistically significant differences for the remaining comparisons (Table 5; Figure 6).

5.4 Discussion

The OSCE has been used for both formative and summative assessments of trainee competence for several years (Pugh et al., 2016). However, there has been a paucity of literature on how OSCE scores relate to clinical performance (Hatala et al., 2015). Of the studies that exist, there was large variability in the skills tested, scales used, and time between assessments (Wagner et al., In preparation). In our study, the OSCE assessments and WBAs were completed by the same population of raters (staff surgeons, fellows, and senior residents), included a common global rating of entrustment, and were completed within one month of each other. Our results suggest that with the exception of suturing, OSCE scores were significantly different than WBA scores. We did not find any statistically significant correlations between OSCE scores and WBA scores, regardless of skill (Table 4). Lastly, formative assessments completed during the BC on suturing were significantly different, and not significantly correlated with WBAs on suturing (Table 5).

5.4.1 Why are OSCE and WBA Scores Different?

There are a number of potential explanations for these results. For all tasks, the mean WBA scores were higher than mean OSCE scores (Table 3). This finding might be a result of time constraints within OSCEs. During the OSCE, residents had two-minutes to read the instructions and eight-minutes to complete the task. It is possible that residents were unable to complete the tasks within the allotted time. However, 10-minute OSCE stations are very common (Khan, Gaunt, Ramachandran, & Pushkar, 2013), and time constraints have not often been cited as a complaint.

Another explanation might be rater training. While the OSCE and WBAs raters were from the same population in our study (staff surgeons, fellows, and senior residents from the McMaster Department of Surgery), the OSCE raters received a brief introduction to the assessment tool, whereas raters in the clinical environment did not. As global ratings of entrustment are still fairly new to medical education, it is possible that WBA raters' unfamiliarity with the scale meant they were still hesitant to provide low ratings (recall the 'Failure to Fail' phenomenon). However, one of the reasons for the movement towards entrustment-based scales is that they are intuitive and do not require rater training (Rekman et al., 2016). Moreover, even if time and/or rater bias resulted in lower OSCE scores/higher WBA scores, neither explanation addresses the lack of correlation between the measures as the Spearman rank correlation looks at rank values rather than absolute scores. Thus, we would still expect someone who scored higher in the OSCE to score higher on WBAs, relative to their peers.

A third explanation might be that the participants included in this study were too homogeneous. Previous research suggests OSCEs and WBAs are able to discriminate between levels of expertise (Pugh et al., 2015). However, all participants in this study were in same year of surgical residency. Accordingly, scores were grouped around 3/5 and 4/5 for both the OSCE and WBAs. With performance between the residents being so similar, it is not surprising that the paired correlations were negligible. Future research should explore the relationship between OSCE scores and WBAs across multiple years of training.

The above notwithstanding, we propose an alternative hypothesis. Previous work suggests that both OSCEs (Pugh et al., 2015) and entrustment-based WBAs (MacEwan, Dudek, Wood, & Gofton, 2016) have fairly robust measurement properties. If both measures are reliable, but do not correlate with one another, we suggest that they might be capturing different aspects of performance.

5.4.2 Miller's Pyramid and Assessment

Miller (1990) proposed that medical trainees develop competence in a stepwise, or graded manner. Miller's educational schema, commonly referred to as Miller's Pyramid, was adapted from Bloom's Taxonomy of learning (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). In Miller's Pyramid, medical trainees move from 'knowing', to 'knowing how', to 'showing how', and to 'doing' over the course of training. To date, many educators have used Miller's Pyramid to discuss assessment strategies in medical training. Educators propose that written examinations provide a way to capture what trainees 'know' (Miller, 1990). Oral examinations provide a way to capture what trainees 'know how' to do (Wass, Van der Vleuten, Shatzer, & Jones, 2001). Simulation-based tests, such as OSCEs, provide a way to capture whether trainees can 'show how' to do something (Miller, 1990). Finally, WBAs provide a way to capture what a trainee 'does' in the clinical environment (Iobst et al., 2010). The results from this study suggest what a trainee 'shows' in a simulated assessment might not always equate to what a trainee 'does' in the clinical environment. They might represent two different aspects, or levels, of competence. Our findings from chapter four support this hypothesis, as residents suggested that OSCEs measured their acting / testing ability rather than true clinical abilities. As a result of these findings, we must consider how we are using OSCEs in medical training.

5.4.3 Implications for Medical Training

To date, OSCEs have been used for many summative licensing examinations around the world (Hodges, 2007). The fact that OSCEs may not represent what a trainee 'does' in the clinical environment raises some questions about the validity of those examinations. With the shift to competency-based medical education (CBME), there has been movement away from summative examinations, towards more frequent formative examinations and portfolios of assessment (chapter 1.7.5). Given our findings, this change in assessment strategies is definitely a step in the right direction; however, there are still some lingering concerns.

As discussed in chapter 1.4, there are a number of new pressures on the healthcare system, which have led to a reduction in the amount of clinical teaching time available to trainees. Due to these new pressures, programs have had to become increasingly reliant on simulation, specifically OSCEs, as an integral component of assessment programs. With the implementation of CBME, and call for more formative assessments, it is likely this reliance will grow. To that end, many specialties in Canada that have transitioned to the new CBME curriculum are allowing trainees to be assessed in simulation *and/or* the clinical environment for certain tasks, or Entrustable Professional Activities (Association of Faculties of Medicine of Canada Working Group, 2016). If WBAs and simulation-based assessments, such as OSCEs are not correlated to one another, we caution educators from using them interchangeably.

5.4.4 Revisiting Assessment in Medical Training

Our findings in chapter four suggest that residents felt OSCE scores did not accurately capture their clinical abilities. However, residents also said that OSCEs provided them with valuable feedback (chapter 4.3.2). Based on our literature review in chapter two, we also know that targeting assessments and feedback to trainees' stage of learning is critical to skill development (chapter 2.3.7). Based off this feedback literature, we suggest that assessments in medical education would be most effective if they followed the framework set out in Miller's education schema (Dauphinee, 1995; Miller, 1990). Trainees should first demonstrate their knowledge on a written test, followed by an oral examination. Following these two assessments, trainees should be afforded the opportunity to practice using simulation, and receive feedback on their performance (perhaps by way of OSCE). Once a trainee has demonstrated sufficient competence in simulation, they should be given the opportunity to work in the clinical environment, and be required to complete a number of WBAs before being awarded a license to practice.

In Canada, trainees complete a written examination at the end of medical school; however, by the time they write their examination they have already spent several years working as a clerk in the clinical environment. Similarly, trainees must complete an OSCE within 12-24 months of graduating from medical school; however, during these 12-24 months they work as a resident in the clinical environment. Moreover, despite the plan to include more frequent WBAs with CBME, trainees still have to complete a written and oral examination at the end of residency to get their license to practice. Thus, while components of Miller's Pyramid are present in medical training, our data suggest that they are currently being implemented somewhat out of order. This is a concern, because we want to ensure the information we are using to make summative decisions is appropriate for the trainee's stage of learning.

We suggest there are a number of reasons that current assessment strategies lack hierarchical alignment with Miller's pyramid. First, there is a need to balance medical education and service provision. Preventing learners from entering the clinical environment until they have passed a written, oral, and simulation-based examination would disrupt clinical services and be extremely costly/resource-intensive. Secondly, while relying on WBAs as a measure of clinical performance is preferred, it is possible that trainees may not have the opportunity to see, or be assessed on some clinical presentations, particularly the rare clinical presentations or conditions. In those cases, assessing performance in simulation is the only option if trainees are to graduate in a timely fashion.

Despite these constraints, we suggest that there is room for improvement in how we assess trainees' performance, in particular in considering how data from different assessment methods are interpreted and utilized. Based on the findings of this study, we suggest OSCEs, and simulation more broadly, can be a valuable tool for providing medical trainees with experience and formative feedback in a safe environment. However, we caution educators from using OSCEs as a summative index of what a trainee 'does' in the clinical environment. As we move towards models of education that become increasingly reliant on frequent assessment, such as CBME, we encourage programs to explore new opportunities to consider this concept of progressive assessment.

5.5 Limitations

As WBAs take time away from clinical responsibilities, we had to be mindful of the number of assessments that we asked trainees to complete. We felt completing six brief assessments, which included at least two different skills over a four-week period was reasonable and allowed for variable clinical experiences while still providing useful data. However, this limited our analyses as we were unable to complete a repeated-measures ANOVA and compare scores across different skills. Secondly, we allowed residents to ask any staff surgeon, fellow, or senior resident from the Department of Surgery at McMaster to complete the WBAs. While having trainees initiate assessments is in line with CBD and other competency frameworks (Wagner et al., 2019), there is concern that this may bias results. Namely, trainees may only ask for assessments from raters known to provide high ratings ("Doves"), avoid raters known to provide low ratings ("Hawks"), and select skills they are comfortable performing (Govaerts, 2006). Thus, it is possible that the higher WBAs scores were due to residents 'cherry-picking' assessment opportunities. Yet, as WBAs are to be resident-driven in the new CBD curriculum, this format is reflective of what happens in practice. Thus, the findings of this study and caution to be mindful of how WBAs and OSCE assessments are being used still holds value. In addition to this, there were some limitations with our secondary analysis. We were able to obtain formative assessments during the suturing BC sessions as they were primarily self-directed practice sessions, with five or six floating facilitators, and occurred three times over the course of the BC. However, it was much more challenging to collect assessments during other BC sessions, which only occurred once during the two weeks, involved more structured didactic components, or had fewer facilitators. As such, we were only able to compare a subset of the OSCE and WBA data to formative assessments completed in simulation. Future research should explore how other formative assessments completed in simulation compare to OSCEs and WBAs. Lastly, while previous literature suggests that both OSCEs and WBAs using entrustment ratings have strong reliability evidence, it is possible that the reliability in our study was limited due to the restricted range of scores (i.e. residents all scoring around 3/5 or 4/5). Future studies should compare OSCE and WBAs across multiple years of training to explore whether the findings from our study are replicated when there is more variability amongst participant scores.

5.6 Conclusions

Objective Structured Clinical Examinations (OSCEs) have been considered the gold standard for assessment in medical education for decades. However, to date there has been little information on how OSCE scores relate to clinical performance. Our study leveraged the implementation of CBME to compare OSCE scores and WBAs scores on five skills, including both technical and non-technical skills. Our results suggest that when evaluating scores on the same skill, with the same measurement scale, and same population of raters within the first month of residency, mean OSCE and WBA scores were significantly different and not significantly correlated. Based on Miller's Pyramid, we propose that these two assessment strategies might be capturing different aspects of a trainees' competence. The OSCE captures what a trainee can 'show how' to do, while WBAs capture what a trainee 'does' in the clinical environment. As our results suggest these two domains are not necessarily related, we propose that OSCEs and other performance-based measures in simulation might have great utility for providing trainees with formative feedback; however we urge educators to carefully consider their appropriateness as a summative index of clinical performance.

Table 1

Participant Demographics

Specialty	Number of Residents	Gender (% Female)	Gender (% Male)
General Surgery	8	6/8 (75%)	2/8 (25%)
Obstetrics and Gynecology	7	6/7 (86%)	1/7 (14%)
Orthopaedics	5	1/5 (20%)	4/5 (80%)
Urology	3	1/3 (33%)	2/3 (67%)
Plastic Surgery	3	1/3 (33%)	2/3 (67%)
Otolaryngology – Head and Neck Surgery	2	1/2 (50%)	1/2 (50%)
Neurosurgery	2	1/2 (50%)	1/2 (50%)
Vascular Surgery	1	0/1 (0%)	1/1 (100%)
Cardiac Surgery	1	1/1 (100%)	0/1 (0%)
TOTALS	**32**	**18/32 (56%)**	**14/32 (44%)**

Table 2

Summary of Assessment Strategies

	Boot Camp Formative Assessments	**Post-Boot Camp OSCE**	**Workplace-Based Assessments**
Assessment Dates	July 3– July 12, 2017	July 12, 2017	July 17– August 16, 2017
Skills Assessed	Suturing	Suturing Breaking bad news Informed consent Foley catheter insertion Hand Ties Recognition of the critically ill patient Pre-operative preparation Handover	Suturing Breaking bad news Informed consent Foley catheter insertion Hand ties
Assessment Format	3 suturing sessions 1 assessment/session	8 skill-based OSCE stations 1 assessment/skill	1 month working in the clinical environment 6 assessments including at least 2 different skills
Assessors	Faculty, fellows, & senior residents	Faculty, fellows, & senior residents	Faculty, fellows, & senior residents
Evaluation Tool	Multiple global rating scales & narrative feedback (Appendix A)	Checklist & multiple global rating scales (Appendix B)	Global rating scale of entrustment & narrative feedback (Appendix C)

Table 3

Descriptive Statistics

Assessment Strategy	**Skill**	**Number of Assessments**	**# of Residents Completing Assessments**	**Mean**	**SD**
Formative Boot Camp Assessments	Suturing	312	34	3.53	0.51
OSCE Post-Boot Camp Assessments	Suturing	34	34	3.91	0.83
	Breaking bad news	34	34	3.09	0.62
	Informed consent	34	34	3.03	0.39
	Foley catheter insertion	34	34	3.73	0.79
	Hand ties	34	34	3.71	0.72
Workplace-based Assessments	Suturing	79	29	4.14	0.58
	Breaking bad news	14	11	4.57	0.50
	Informed consent	37	22	4.25	0.54
	Foley catheter insertion	20	12	4.48	0.57
	Hand ties	45	26	4.12	0.62

Table 4

Comparison of Means and Correlation Coefficients, by Skill and Assessment Strategy

Assessment Strategy	Skill	# of Residents Included in Comparison	Z - value	*p*	Correlation Coefficient	*p*	Bootstrapped 95% CI
OSCE vs. Workplace-based Assessments	Suturing	29	1.42	.16	$r_s = .13$	.51	- .25, .50
	Breaking bad news	11	2.99	**< .01**	$r_s = .28$	.39	- .34, .73
	Informed consent	22	4.05	**< .01**	$r_s =.28$	.21	- .05, .55
	Foley catheter insertion	12	2.55	**.01**	$r_s = .16$	.62	- .53, .78
	Hand ties	26	2.03	.04	$r_s = .16$	.42	- .21, .50

Note: Statistical significance considered when p < .01

Table 5

Suturing Scores Across Assessment Strategies: ANOVA Results

	Comparison	Test Statistic	*df*	*p*	Correlation Coefficient	*p*
Main Effects	Assessment strategy	F = 15.77	2, 56	**< .01**		
Pairwise Comparisons	BC Formative vs. OSCE	t = - 2.23	28	.10	r_s = .27	.12
	OSCE vs. Workplace-based	t = - 1.35	28	.57	r_s = .13	.51
	BC Formative vs. Workplace-based	t = - 3.97	28	**< .01**	r_s = -.11	.57

Note: Statistical significance considered when p < .05 for the ANOVA. Statistical significance considered when p < .02 for the pairwise correlations.

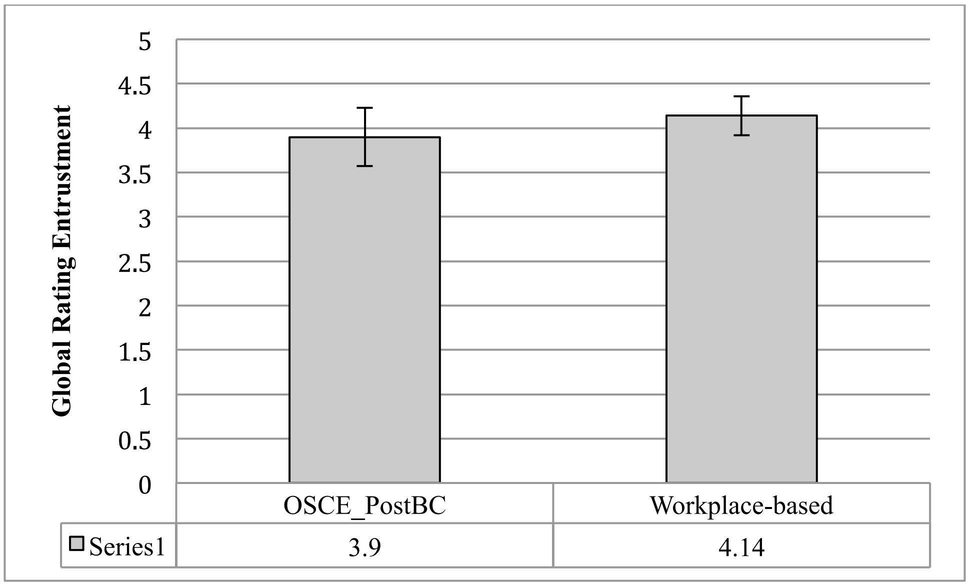


Figure 1. Post-Boot Camp OSCE and Workplace-based Assessments (Suturing)

Error bars show 95% confidence interval

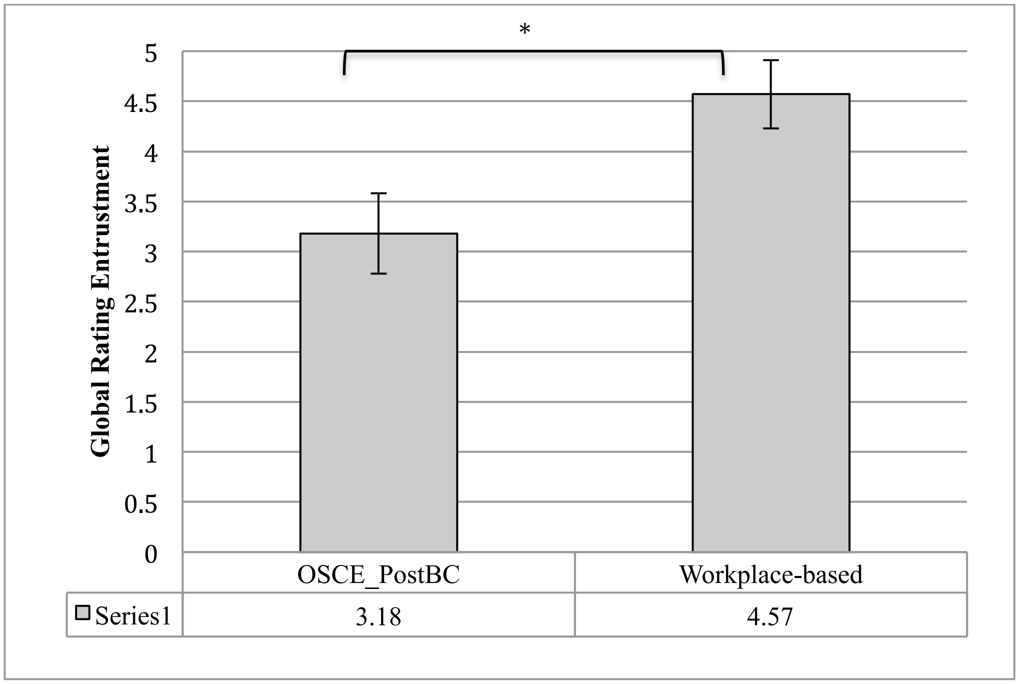


Figure 2. Post-Boot Camp OSCE and Workplace-Based Assessments (Breaking Bad News)

* = Statistically significant difference between assessment scores
Error bars show 95% confidence interval

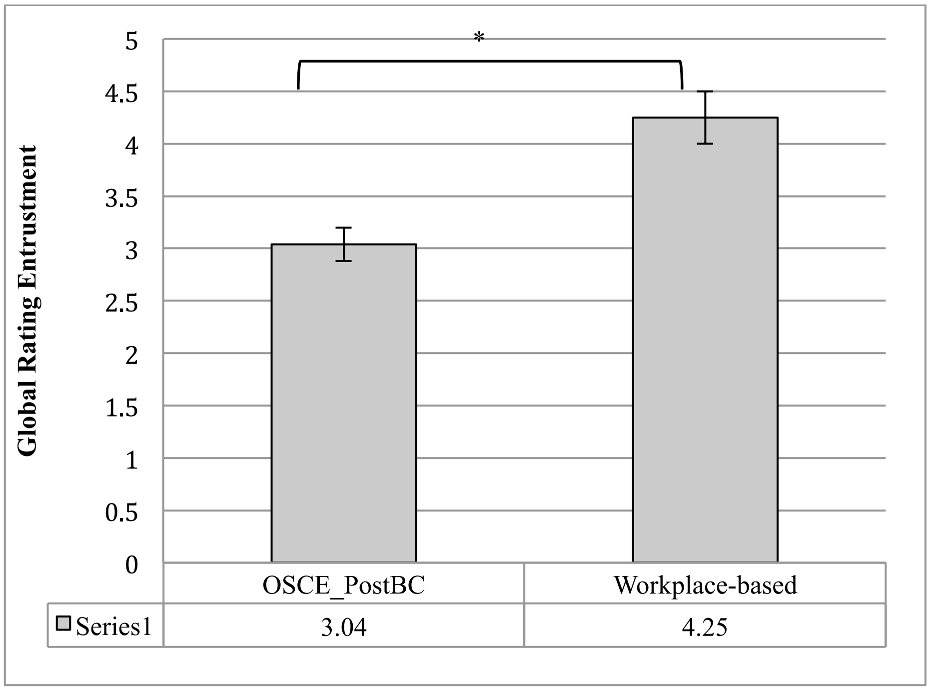


Figure 3. Post-Boot Camp OSCE and Workplace-based Assessments (Informed Consent)

* = Statistically significant difference between assessment scores
Error bars show 95% confidence interval

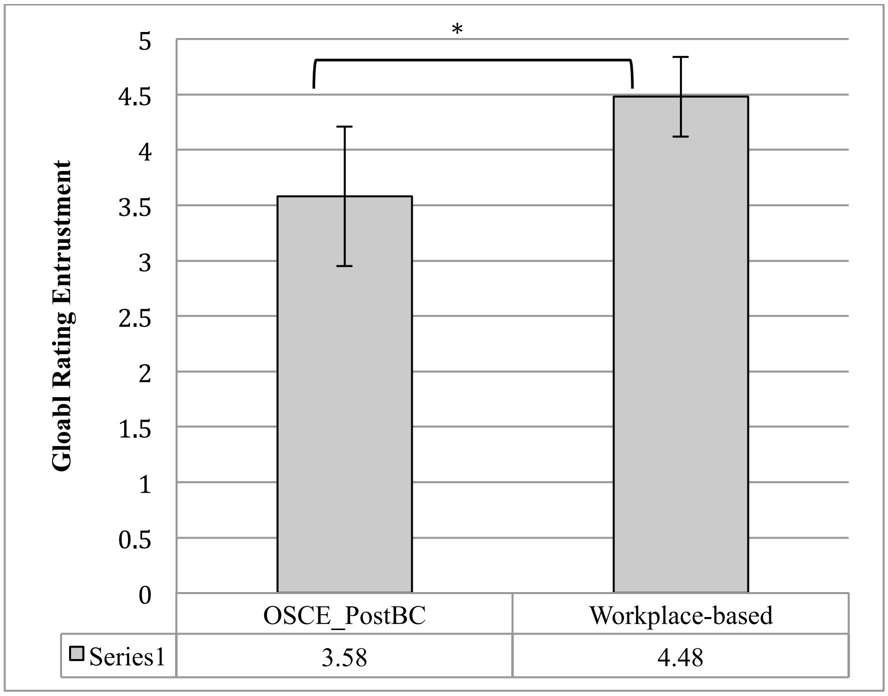


Figure 4. Post-Boot Camp OSCE and Workplace-based Assessments (Foley Catheter Insertion)

* = Statistically significant difference between assessment scores
Error bars show 95% confidence interval

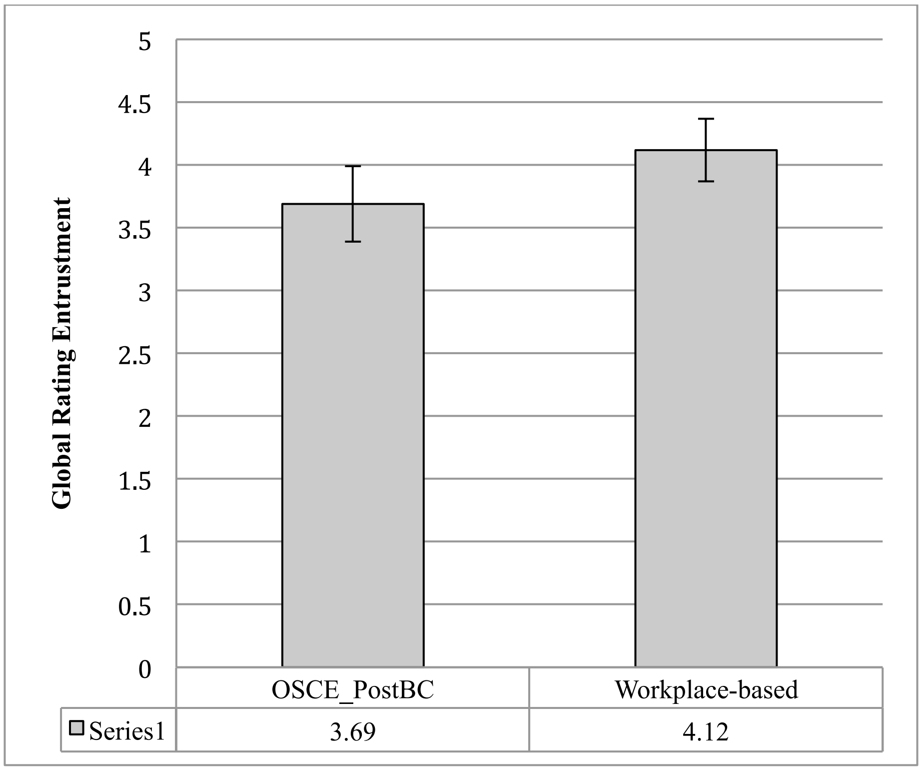


Figure 5. Post-Boot Camp OSCE and Workplace-based Assessments (Hand Ties)

Error bars show 95% confidence interval

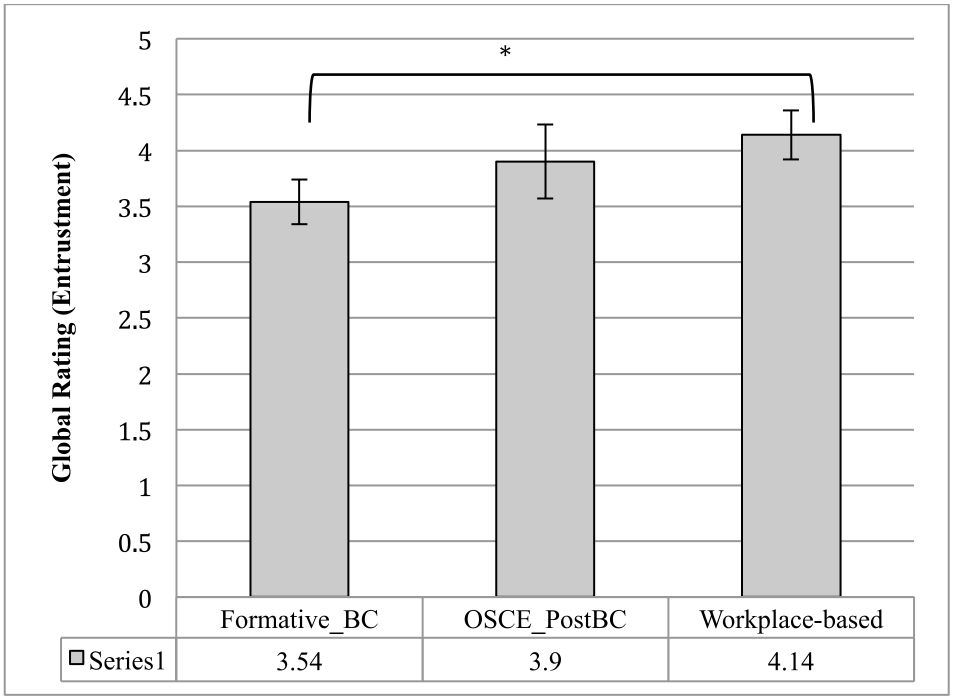


Figure 6. Suturing Scores Across Assessment Strategies

* = Statistically significant difference between assessment scores

Error bars show 95% confidence interval

Appendix A

McMaster Surgical Foundations Formative Boot Camp Assessment

Date	
Resident	
Evaluator	
Suturing Technique Demonstrated	

Please circle the number corresponding to the candidate's performance observed today, regardless of their level of training.

Time and motion

1	2	3	4	5
Slow and made many unnecessary movements		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency

Instrument handling

1	2	3	4	5
Repeatedly made tentative or awkward movements		Occasionally appeared stiff or awkward		Fluid movements with no stiffness or awkwardness

Respect for Tissue

1	2	3	4	5
Frequently used unnecessary force on tissue/caused damage by inappropriate use of instruments		Careful handling of tissue by occasionally caused inadvertent damage		Consistently handled tissue appropriately; caused minimal damage to tissue

Entrustment

1	2	3	4	5
Requires complete supervision/guidance		Requires some supervision/prompting		Requires no supervision; ready for independent practice

Please give one suggestion for improvement:

Appendix B

Sample OSCE Assessment Form

Instructions to Candidates:
Demonstrate the following stitches with instrument ties:
- 2 horizontal mattress sutures
- 2 vertical mattress sutures
- 1 running subcuticular suture

ITEM	Not done, Incorrect	Done, Correct
1. Successfully completed 2 horizontal mattress sutures	0	1
2. Successfully completed 2 vertical mattress sutures	0	1
3. Successfully completed 1 running subcuticular suture	0	1
TOTAL SCORE		**/3**

Time and motion

1	2	3	4	5
Slow and made many unnecessary movements		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency

Instrument handling

1	2	3	4	5
Repeatedly made tentative or awkward movements		Occasionally appeared stiff or awkward		Fluid movements with no stiffness or awkwardness

Respect for Tissue

1	2	3	4	5
Frequently used unnecessary force on tissue/caused damage by inappropriate use of instruments		Careful handling of tissue by occasionally caused inadvertent damage		Consistently handled tissue appropriately; caused minimal damage to tissue

Entrustment

1	2	3	4	5
Requires complete supervision/guidance		Requires some supervision/prompting		Requires no supervision; ready for independent practice

Overall Performance

1	2	3	4	5
Demonstrated no competence; requires extra support		Demonstrated some competence; remains in training		Demonstrates complete competence; ready for independent practice

Appendix C

McMaster Surgical Foundations Workplace-Based Assessment Form

Resident Name:	**Date:**
Assessor Name:	**Assessor Role (please circle):** Senior Resident Fellow

Task (please circle task evaluated):

Obtaining Informed Consent Breaking Bad News Foley Catheter Insertion Hand Ties Suturing

Entrustment Rating (please circle the number corresponding the performance <u>observed today</u>)

1	2	3	4	5
Requires complete guidance/supervision		Requires some guidance/supervision		Requires no supervision; ready for independent

Please give one suggestion for improvement:

NOTES:

- Residents are required to complete **6 formative assessments**
- The 6 assessments must include **at least 2 different tasks**
- Completed evaluations are due **Wednesday August 16th @ 9am** to the Surgical Foundations Program Coordinator

5.7 References

Association of Faculties of Medicine of Canada Working Group. (2016). *The Association of Faculties of Medicine of Canada Entrustable Professional Activities for the Transition from Medical School to Residency*. Retrieved from https://afmc.ca/sites/default/files/documents/AFMC_Entrustable_Professional_Activities_EN_0.pdf

Beard, J., & Bussey, M. (2007). The ISCP Learning Portfolio. *Ann R Coll Surg Engl, 89*, 200-202. doi:10.1308/147363507X204486

Blanca, M. J., Alarcon, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552-557. doi:10.7334/psicothema2016.383

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: David McKay Company.

Bridge, P. D., & Sawilowsky, S. S. (1999). Increasing Physicians' Awareness of the Impact of Statistics on Research Outcomes: Comparative Power of the t-test and Wilcoxon Rank-Sum Test in Small Samples Applied Research. *J Clin Epidemiol, 52*, 229-235. Retrieved from https://journals-scholarsportal-info.libaccess.lib.mcmaster.ca/pdf/08954356/v52i0003/229_ipaotiosoro.xml.

Charvat, J., McGuire, C., & Parsons, V. (1968). A review of the Nature and Uses of Examinations in Medical Education. *Public Health Papers, 36*.

Cowles, J. T. (1954). Current Trends in Examination Procedures. *JAMA, 155*, 1383-1387.

Dauphinee. (1995). Assessing Clinical Performance: Where Do We Stand and What Might We Expect? *JAMA, 274*, 741 - 743. Retrieved from https://mcc.ca/media/Dauphinee_Assessing-clinical-performance-Where-do-we-stand-and-what-might-we-expect.pdf.

Dudek, N., Gofton, W., Rekman, J., & McDougall, A. (2019). Faculty and Resident Perspectives on Using Entrustment Anchors for Workplace-Based Assessment. *Journal of Graduate Medical Education*. doi:10.4300/jgme-d-18-01003.1

Dudek, N., Marks, M. B., & Regehr, G. (2005). Failure to fail: The perspectives of clinical supervisors. *Acad Med, 80*(10), S84-S87. Retrieved from https://ovidsp-tx-ovid-com.libaccess.lib.mcmaster.ca/.

Govaerts, M. J. (2006). In-training assessment: learning from practice. *The Clinical Teacher, 3*, 242-247.

Govaerts, M. J., Van de Wiel, M. W., Schuwirth, L. W., Van der Vleuten, C. P., & Muijtjens, A. M. (2013). Workplace-based assessment: raters' performance theories and constructs. *Adv Health Sci Educ Theory Pract, 18*(3), 375-396. doi:10.1007/s10459-012-9376-x

Grand'Maison, P., Lescop, J., & Brailovsky, C. A. (1993). Canadian experience with structured clinical examinations. *Can Med Assoc J, 148*, 1573-1576.

Harden, R. M., Stevenson, M., Wilson Downie, W., & Wilson, G. M. (1975). Assessment of Clinical Competence using Objective Structured Examination. *British Medical Journal, 1*, 447-451. Retrieved from https://www-jstor-org/.

Hart, I. (1980). The use of the objective structured clinical examination to compare clinical skills in two medical schools. *Annals of the Royal College of Physicians and Surgeons of Canada, 13*, 145.

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. *Adv Health Sci Educ Theory Pract, 20*(5), 1149-1175. doi:10.1007/s10459-015-9593-1

Hodges, B. (2007). *A socio-historical study of the birth and adoption of the objective structured clinical examination (OSCE).* (Doctor of Philosophy), University of Toronto, Canada.

Hodges, B., Turnbull, J., Cohen, R., Bienenstock, A., & Norman, G. (1996). Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. *Medical Education, 30*, 38-43. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8736187.

Iobst, W. F., Sherbino, J., Cate, O. T., Richardson, D. L., Dath, D., Swing, S. R., . . . Frank, J. R. (2010). Competency-based medical education in postgraduate medical education. *Med Teach, 32*(8), 651-656. doi:10.3109/0142159X.2010.500709

Khan, K. Z., Gaunt, K., Ramachandran, S., & Pushkar, P. (2013). The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part II: organisation & administration. *Med Teach, 35*(9), e1447-1463. doi:10.3109/0142159X.2013.818635

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med, 15*(2), 155-163. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27330520. doi:10.1016/j.jcm.2016.02.012

MacEwan, M. J., Dudek, N. L., Wood, T. J., & Gofton, W. T. (2016). Continued Validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): Use in the Simulated Environment. *Teach Learn Med, 28*(1), 72-79. doi:10.1080/10401334.2015.1107483

Massie, J., & Ali, J. M. (2016). Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Adv Health Sci Educ Theory Pract, 21*(2), 455-473. doi:10.1007/s10459-015-9614-0

McGuire, C. (1966). The Oral Examination as a Measure of Professional Competence. *Journal of Medical Education, 41*, 267-274.

McQueen, S. A., Petrisor, B., Bhandari, M., Fahim, C., McKinnon, V., & Sonnadara, R. R. (2016). Examining the barriers to meaningful assessment and feedback in medical training. *Am J Surg, 211*(2), 464-475. doi:10.1016/j.amjsurg.2015.10.002

Miller, G. E. (1990). The Assessment of Clinical Skills/Competence/Performance. *Acad Med, 65*, s63 - s67. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2400509.

Norcini, J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach, 29*(9), 855-871. doi:10.1080/01421590701775453

Pugh, D., Hamstra, S. J., Wood, T. J., Humphrey-Murto, S., Touchie, C., Yudkowsky, R., & Bordage, G. (2015). A procedural skills OSCE: assessing technical and non-technical skills of internal medicine residents. *Adv Health Sci Educ Theory Pract, 20*(1), 85-100. doi:10.1007/s10459-014-9512-x

Pugh, D., Touchie, C., Humphrey-Murto, S., & Wood, T. J. (2016). The OSCE progress test--Measuring clinical skill development over residency training. *Med Teach, 38*(2), 168-173. doi:10.3109/0142159X.2015.1029895

Regehr, G., Macrae, H., Reznick, R., & Szalay, D. (1996). Comparing the Psychometric Properties of Checklists and Global Rating Scales for Assessing Performance on an OSCE-format Examination. *Acad Med, 73*, 993- 997.

Rekman, J., Gofton, W., Dudek, N., Gofton, T., & Hamstra, S. J. (2016). Entrustability Scales: Outlining Their Usefulness for Competency-Based Clinical Assessment. *Acad Med, 91*(2), 186-190. doi:10.1097/ACM.0000000000001045

Reznick, R., Smee, S., Rothman, A., Chalmers, A., Swanson, D., Dufresne, L., . . . Berard, M. (1992). An Objective Structured Clinical Examination for the Licentiate: report of the Pilot Project of the Medical Council of Canada. *Acad Med, 67*, 487-494. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1497774.

van der Vleuten, C. P., Schuwirth, L. W., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Med Teach, 34*(3), 205-214. doi:10.3109/0142159X.2012.652239

Wagner, N., Acai, A., McQueen, S. A., McCarthy, C., McGuire, A., Petrisor, B., & Sonnadara, R. R. (2019). Enhancing Formative Feedback in Orthopaedic Training: Development and Implementation of a Competency-Based Assessment Framework. *J Surg Educ*. doi:10.1016/j.jsurg.2019.03.015

Wagner, N., Fahim, C., & Sonnadara, R. (In preparation). *A systematic review of concurrent validity evidence for objective structured clinical exams*.

Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet, 357*(9260), 945-949. doi:10.1016/s0140-6736(00)04221-5

Chapter 6: General Discussion and Conclusions

6.0 General Discussion and Conclusions

6.1 Preface

Chapter one began by describing the evolution of medical training. From the formation of the College of Physicians and Guild of Barber-Surgeons in London, to the development of medical schools in America and the push for more standardized curriculum, medical training has gone through a number of significant reforms through the ages. As we move into the 21st century, how medical practitioners are trained is once again changing. As the population continues to expand at rapid rates, with people living longer than ever before, healthcare needs are becoming more complex (British Medical Association, 2018). In trying to meet these growing demands, it has become increasingly difficult to balance service provision with trainee education (Qureshi & Maxwell, 2012). Moreover, with reductions to trainee work hours, educators are concerned that time-based models of medical training can no longer ensure that 'x' amount of time can provide trainees with the knowledge and skills required for independent practice (R. Sonnadara et al., 2013). In response to these new pressures, we are now seeing a global shift towards competency-based models, which are based on what a trainee is able to do, rather than time spent training (ten Cate, 2017). As this new framework gains traction, gaps in previous training models, such as the way learners are supported through transitions to new roles, have been revealed.

Throughout their careers, physicians must transition into a number of different roles. However, the transition from medical student to resident is often considered the most challenging year for learners (Bligh, 2002). Starting on day one of residency, trainees are no longer students, but are resident *doctors.* This new title comes with an increase in expectations from both patients and other healthcare providers. New doctors are expected to adapt to new hospitals and training environments, manage increased workloads, and make decisions about patient care, while working longer hours, with less support than they had as medical students. This increase in responsibility would be difficult at any point, but it is amplified by the fact that trainees typically have several months off between their last clinical placement and the first day of residency. The break from clinical placement is to allow students to study for the written examination (the first step of licensing examinations in Canada). Yet having this break means that, in addition to adjusting to a new environment and role, new doctors must reacquaint themselves with the procedures and responsibilities of working in the clinical environment. Depending on trainees' clerkship rotation schedule and which specialty they picked to pursue, it can be as much as six to eight months since they were in a similar clinical environment. Thus, despite the increased responsibilities at the onset of residency, new doctors might actually be less competent, or confident, than they were at the end of medical school. As such, literature suggests this transition period often makes trainees feel underprepared and stressed (Cave, Woolf, Jones, & Dacre, 2009), and raises concerns about patient safety (Young et al., 2011).

In 2010, Phillips and Barker conducted a retrospective study of death certificates in the United States. Through this study, Phillips and Barker (2010) identified a decrease in healthcare efficiency and patient safety, with fatal medical errors rising by 10% in July compared to all other months. As this increase in fatal errors was only seen in counties with teaching hospitals, Phillips and Barker concluded that the "July Effect" was due (at least in part) to changes associated with the arrival of new medical residents. This study was not the only one to point to the idea of increased complications at the beginning of the academic year. Studies in the United Kingdom, where the academic year for medical trainees starts on the first Wednesday in August rather than July 1st, have found similar findings. For example, Jen, Bottle, Majeed, Bell, and Aylin (2009) conducted a retrospective study on patients admitted to hospitals in England the last week of July versus the first week of August from 2000 to 2008. Their results suggest that patients admitted in the first week in August (when trainees in England start their new roles) had a 6-8% increased mortality rate compared with those admitted in the last week of July when controlling for year, gender, age, socio-economic status and other comorbidities (Jen et al., 2009). Moreover, a report by Dr. Foster Intelligence in the United Kingdom provided further support for 'Black Wednesday', as they evaluated 300,000 emergency admissions over nine years and once again found a 6-8% increase in mortality in the first week of August compared to the last week of July (Gaskell, Hinton, Page, Elvins, & Malin, 2016; Imperial College of London, 2009). Given that trainees report feeling underprepared and stressed, and there is evidence to suggest that this lack of preparedness may negatively impact patient safety, the transition from medical school to residency is a critical point in medical education in which learners need to be better supported.

6.2 Thesis Overview

In this thesis, we sought: 1) to improve the way that surgical training programs approach the transition into residency in a way that is embedded in learning theory and aligns with the most recent shift towards competency-based models of medical education; and (2) to look at the longer-term impacts of these changes. Initially, a literature review explored principles of novice skill acquisition from the cognitive science, education science and motor behavior perspectives and how those principles can be applied to medical training. Based on the findings of this review, and an understanding of how medical training has evolved over time, we designed three interrelated studies. Study 1 used a Context, Input, Process, and Product (CIPP) framework to develop, implement, and evaluate short-term outcomes of a two-week, simulation-based boot camp (BC) for novice surgical trainees, which was intended to help prepare them for their new role as residents. Study 2 then used a Convergent Parallel Mixed Methods (CPMM) approach to explore longer-term effectiveness of the BC program. Lastly, study 3 explored how effectively the Objective Structured Clinical Examination (OSCE), a "gold standard" measure of learner competency that was used in the BC program, truly captures clinical performance of novice trainees. The key findings from this thesis are summarized below.

6.3 Key Findings

6.3.1 Literature Review on the Theoretical Principles of Novice Skill Acquisition

The initial literature review was divided into three sections. The first section explored theoretical principles of novice skill acquisition and encompassed topics such as: motor programs, Hebbian learning and long-term potentiation, stages of motor learning, working memory, cognitive load theory, divided attention, and dual-task conditions. The second section investigated how best to structure practice sessions to optimize learning in novice learners, including topics such as deliberate practice, observational learning, practice schedules and contextual interference, assessment, and feedback. The third and final section examined how these concepts might be incorporated into the early stages of postgraduate medical training, specifically a BC at the onset of residency.

6.3.2 Study 1: Development, Implementation and Evaluation of a BC for Novice Trainees

Based on the findings from the literature review, we used the Context, Input, Process, and Product (CIPP) program evaluation model to develop, implement and evaluate the short-term outcomes of a BC for novice surgical trainees. Through the Context evaluation phase, we found that both faculty and residents at McMaster agreed with previous literature suggesting there is a need to support trainees more effectively through the transition from medical school to residency. During the Input evaluation phase, BC curriculum was designed based on the recommendations from our literature review on novice skill acquisition, guidelines from the Royal College of Physicians and Surgeons of Canada (RCPSC), as well as input from faculty and residents at McMaster. In the Process evaluation phase, daily observational notes and exit interviews with facilitators suggested the BC was implemented, for the most part, according to the intended plan. During the Product evaluation phase, we identified that the BC significantly improved residents' self-rated preparedness for residency. Moreover, the majority of residents (90%) felt the BC was extremely useful in preparing them for residency. Through this evaluation as well as session-specific feedback, we concluded that the BC was successful in meeting its educational goals.

This study also allowed us to identify a number of themes across the CIPP phases. While resident self-rated preparedness scores increased following the BC, some residents still felt more prepared than others. Participants suggested this variability reflected the fact that individuals graduate medical school with different clinical experiences. Within the same medical school, learners have different experiences due to the unpredictable nature of what comes into the hospital or clinic on any given day. Between medical schools, learners have different experiences due to the size of the program/number of other learners, exposure to different rotations/specialties, and different expectations about the role of medical students in the clinical environment. These experiences play a major role in what new residents are comfortable dealing with. Despite the variability in starting points across those participating in the BC, participants felt the BC was extremely useful to review expectations and 'make sure everyone was on the same page', helped them feel more confident and safer providing patient care, provided them with an opportunity to meet other residents and staff surgeons, and to become familiar with their new environment.

6.3.3 Study 2: Long-Term Impacts of a Boot Camp for Novice Trainees

The second study used a Convergent Parallel Mixed Methods (CPMM) approach to explore longer-term impacts of our BC program. Quantitative results that explored performance on various targeted skills suggested that, on average, the BC cohort scored higher on the OSCE than the Traditionally Trained (TT) cohort at both one- and two-years into training. However, consistent with previous literature (e.g. R. R. Sonnadara et al., 2012), the differences between the two cohorts were station-specific. Meanwhile, qualitative findings suggested that residents felt BCs provided an opportunity to become acquainted with hospital logistics and to review basic clinical skills and expectations. Program Directors echoed this sentiment and suggested that BCs were a way to efficiently deliver information to trainees and 'level the playing field'. Both residents and Program Directors reported that the social component of the BC was extremely valuable, as residents were able to meet their peers, senior residents, and faculty in a low-stress environment. Residents felt this provided them with a social support system and set the tone for maintaining good relationships throughout the remainder of their training program. Lastly, both residents and Program Directors suggested that while BCs might not fully prepare them for residency, the BC provided them with some basic knowledge, confidence, and a support system to handle clinical situations. As such, the majority of participants felt the BC was a useful educational program and should be rerun for all incoming cohorts.

By using a mixed methods approach to combine our quantitative and qualitative findings, we were also able to use session-specific feedback to provide context for some of the OSCE scores. This analysis revealed that residents felt that some BC sessions were more useful than others, mainly those that included 'practical information' or 'hands-on' practice sessions. While this feedback explained some OSCE results, residents also stated that some OSCE stations lacked fidelity, or did not 'feel' realistic. Consequently, residents felt the OSCE might not be reflective of their true clinical abilities. Based on this finding, plus the fact that the OSCEs failed to capture some of the perceived benefits of the BC, such as refreshing clinical knowledge, reviewing expectations and hospital logistics, and meeting other residents, we concluded that OSCEs alone, might not be the best measure of long-term BC effectiveness.

6.3.4 Study 3: OSCE Assessments as an Indicator of Clinical Performance

In the final study, we explored the relationship between OSCEs and clinical performance of novice trainees, measured through Workplace-based Assessments (WBAs). Results from this study suggested that OSCE scores and WBA scores were indeed significantly different and not correlated. As these assessments were completed on the same group of residents, rated by the same population of raters, on the same skills, using the same scale, all within one month of each other, we suggest that OSCEs and WBAs might capture different aspects of performance. Based on these results, we suggest that educators should carefully consider how they are using simulation-based assessments, such as OSCEs, in training.

6.4 Discussion

6.4.1 Selecting Appropriate Methods

Previous work suggests that BCs at the onset of residency can be an effective way to improve trainees' skills and privilege later clinical learning (R. R. Sonnadara et al., 2013; R. R. Sonnadara et al., 2012; R.R Sonnadara et al., 2011). Our work supports these findings, as the BC cohort scored higher than the Traditionally Trained (TT) cohort at both one and two years into training, and the BC cohorts' scores continued to improve from post-BC to two years into training. In addition to reaffirming these results, our approach uncovered many new insights regarding BCs.

To the best of our knowledge, this thesis was the first to use a formal program evaluation method, specifically the Context, Input, Process, and Product (CIPP) model, to study the development, implementation, and early outcomes of a BC program. Using the CIPP approach allowed us to create a BC based in learning theory, while also targeting the new RCPSC transition to discipline curriculum, and local educational needs. Moreover, by documenting our approach, we were able to make informed changes to future iterations of the BC at McMaster, while also providing educators with meaningful information on why certain decisions were made, allowing others to reproduce aspects of the BC that are applicable to them, and modify where needed.

Similarly, this thesis was the first to use a Convergent Parallel Mixed Methods (CPMM) approach to study long-term effectiveness of a BC. With OSCEs at one and two years into training, this body of work includes the longest follow-up period in the BC literature. As we know it can take several years to identify the intended and unintended consequences of an educational intervention (Bland et al., 2000), this extended timeframe provides critical information about the efficacy of BCs. At a local level, our approach has resulted in the BC being formally adopted by the McMaster Department of Surgery as part of the curriculum for all incoming trainees. More broadly, we hope that this long-term evidence might give educators the 'proof' they need to inspire change at their own institutions and clarify the value of this kind of educational intervention. Mixed methods are not often used in medical education research, as they can be both time- and resource-intensive, and require a familiarity with qualitative, quantitative, and mixed methods literature. However, given the importance of context in medical training, we suggest that mixed methods could be extremely valuable for many research questions where qualitative or quantitative methods alone cannot comprehensively address the question. In our case, we were able to use the CPMM approach to link how BC implementation may have affected OSCE scores. Our mixed methods approach also allowed us to identify many critical themes, which demonstrated that the benefits to BCs extend far beyond direct improvements to resident skill level.

6.4.2 Organizational Socialization

Throughout the CIPP and CPMM studies, participants commented on the fact that learners experience the transition to residency in different ways. Some learners feel more prepared than others. Often, participants suggested the varied feelings of preparedness were a result of medical students graduating medical school with different clinical experiences, as well as the length of time since their last surgical clerkship rotation. Many trainees felt they were rusty. Thus, while participants suggested the BC could never truly prepare them for residency, they felt the BC was useful in refreshing basic clinical knowledge and skills; reviewing expectations of new residents; ensuring everyone was on the same page; and providing opportunities to create a social support network. Based on these findings, we suggested that BCs act as a form of Organizational Socialization, or the process of learning the behaviours, attitudes, and social etiquette needed to assume a role within an established organization (Van Maanen & Schein, 1979). The four main components of Organizational Socialization, task mastery, role clarification, acculturation and social integration, are clearly reflected in our findings. We know that Organizational Socialization is critical for anyone entering a transition period (Morrison, 1993); yet, it is something that is rarely talked about in medical training. We hope this work provides a framework for how doctors, and other healthcare professionals, can be supported through a number of other transitions over the course of their careers.

6.4.3 Rethinking Assessment in Medical Education

Objective Structured Clinical Examinations (OSCEs) are largely considered the gold standard of assessment in medical education, and as such, have been used for much of the BC research to date. However, in our study the OSCE failed to capture the importance of Organizational Socialization in BCs. Moreover, our findings from study 2 (the CPMM study) suggested that OSCEs might not be an accurate representation of clinical knowledge. Residents felt that the OSCE was not 'realistic' and Program Directors were unable to identify a difference in clinical performance between the BC and TT cohort at six months into training, despite significant differences in OSCE scores. When we looked at this issue in more detail in study 3, we found that OSCE scores and Workplace-based Assessment (WBA) scores were significantly different and not significantly correlated. Moreover, there was no correlation between formative assessments completed during the BC and WBAs. As we were comparing resident performance on the same scales, same tasks, with the same raters, all within the first month of training, we suggested that performance-based assessments in simulation might be capturing something different than WBAs completed in the clinical environment. This observation was in line with our findings from study 2, but also had support from previous literature, such as Miller's educational schema, which suggests learners 'showing how' to do something in simulation represents a different level of competence than what they can 'do' in the actual clinical environment (Miller, 1990). As performance-based tests, such as OSCEs, are currently being used interchangeably with WBAs in new competency-based frameworks, as well as summative licensing examinations, our findings raise concerns about the validity of current assessment approaches in medical training. While additional research is needed to gather data on how performance-based tests relate to clinical performance for other tasks and across several years of training, these findings, paired with the mixed methods results, start a really interesting conversation about how we are using simulation-based assessments, such as OSCEs, in medical training, and whether that is truly appropriate.

6.4.4 Limitations and Benefits of Applied Research in Medical Education

There are a number of limitations to doing applied research in medical education. First, residents are both learners and working doctors. With increasing pressures on the healthcare system and a reduction to trainee work hours, residents are expected to see more patients in less time. As such, medical training programs are always trying to balance service provision and education. Time spent participating in research, even if it has the potential to improve training, is time spent away from patient care and clinical teaching. Furthermore, while many studies rely on randomized controlled trials, and comparisons with control groups, traditional 'experimental' designs are often not appropriate for medical education research. If there is an approach we think will be more effective, we have a moral imperative to provide it to all learners. This is particularly true when we believe it might ultimately affect patient care. For this reason, we often see studies, including the ones in this thesis, which rely on comparisons with historical controls, or previous cohorts of residents.

Beyond challenges finding appropriate control groups, there are other factors in medical education research that are difficult to control. If studies wish to evaluate performance on some clinical task, they need to rely on expert raters. As expert raters are typically working clinicians, studies must rely on volunteers who participate around their clinical duties. This means that finding consistent raters across all learners is often impossible. Moreover, residents have different clinical experiences across their training. Thus, attributing any performance differences on a task to a specific intervention is very difficult, as is finding appropriate, comparable clinical experiences on which performance can be measured between learners for a specific skill.

Lastly, as learners are expected to stay up to date with current literature and practices, and required to do most, if not all, of their studying outside of clinical work hours, medical education research is unable to control for study habits and independent learning. In our study we ensured there were no curriculum changes, besides the BC, for the 2015 Traditionally Trained (TT) and 2016 (BC) cohort of Surgical Foundations (SF) residents. However, we were unable to control for learning that happened outside of formal teaching sessions.

Often, there is temptation to try to control a lot of these factors in the quest for scientific rigour. However, such rigour brings a heavy price. Without the richness of the environmental context, we cannot accurately capture the issues, study how to solve them, or know that any 'improvement' will actually be effective or practical. While applied research such as the work presented in this thesis has a number of limitations, the strong ecological validity it enables is vital if we want to introduce meaningful, sustainable improvements to medical education.

6.5 Conclusions

Reflecting back on the days of Barber-Surgeons and healers, medical education has come a long way. However, this evolution has not always been straightforward or direct. Medical training is complex, and it is difficult to meet the needs of society and the needs of the learners, especially as those needs evolve over time. These dynamic and complex relationships make medical training difficult to study, and make it even more difficult to implement change. Yet, it is so very critical. As Abraham Flexner wrote: "… Not only the personal wellbeing of each citizen, but the national, state, and municipal sanitation rests upon the quality of the training which the medical graduate has received." (Flexner, 1910, p. XV).

At the local level, this thesis leveraged the transition to competency-based medical education frameworks to implement a boot camp (BC) for incoming surgical trainees at McMaster University. From planning to implementation, the BC initiative included 38 first-year residents, 61 volunteer facilitators, nine graduate and undergraduate research students, and two research assistants, across many different Departments and Faculties. This diverse set of contributors highlights not only how expansive the project was, but also the number of people it impacted. This thesis was able to bring together individuals from different fields who do not often work together, and have them work towards a common goal, and some of these collaborations have now extended to other work. Due to the new research this project has enabled, and the positive responses from residents, faculty, and Program Directors, the Department of Surgery at McMaster has decided to continue implementing the BC for all incoming surgical residents for the foreseeable future. The fact that this BC will live on beyond this thesis work and continue supporting trainees through the transition from medical school to residency reinforces the importance of applied research, and the importance of this work at McMaster.

Within a broader context, this thesis has been able to provide surgical educators with valuable information on how novices develop new skills, principles of novice skill acquisition such as deliberate practice, distributed schedules, observational learning, and formative feedback, and how to incorporate those principles into a BC at the onset of residency. Having this process documented through a CIPP approach provides detailed information on why certain decisions were made, and will allow educators to reproduce aspects of the BC that are applicable to them, and modify where needed, within the context of their own training programs. Additionally, using CPMM to measure long-term BC effectiveness highlighted a new approach for evaluating the efficacy of educational initiatives, and opened a conversation about the need to revisit assessment in medical training, and ensure we are using simulation-based assessments, such as OSCEs, appropriately.

6.6 Epilogue

> *A transition is not a moment, but rather a dynamic process where an individual moves from one set of circumstances to another… In coping with major changes, individuals face new challenges, opportunities, stress and a range of emotions. (Teunissen & Westerman, 2011, p. 52)*

At the start of this thesis, we presented this quote by Teunissen and Westerman (2011) on the nature of transitions. Whether at the level of a major reform, the transition from medical student to resident, or moving assessments from simulation to clinical practice, transitions are dynamic and complex. Yet, moving through any transition creates numerous opportunities.

In our case, we were able to use the transition to competency-based medical education to draw attention to the transition between medical school and residency, implement a BC, and then use that BC to explore how we assess novice trainees. By employing methods that were created for complex and dynamic problems such as CIPP and CPMM, we were able to showcase a number of the intricacies of medical education and provide strong ecological validity while conducting rigorous research that was built upon a robust theoretical framework. Beyond improving our understanding of BCs and assessment, we hope this thesis sheds light on alternative approaches to studying research questions in medical education. Moreover, we hope this thesis sheds light on that fact that transitions, while challenging, are not all bad. Rather, they are powerful in their ability to inspire change.

6.7 References

Bland, C., Starnaman, S., Wersal, L., Moorhead-Rosenberg, L., Zonia, S., & Henry, R. (2000). Curricular Change in Medical Schools: How to Suceed. *Acad Med, 75*, 575 - 594. Retrieved from http://ovidsp.dc1.ovid.com.libaccess.lib.mcmaster.ca/.

Bligh, J. (2002). The first year of doctoring: still a survival exercise. *Medical Education, 36*, 2-3. doi:10.1046/j.1365-2923.2002.01129.x

Cave, J., Woolf, K., Jones, A., & Dacre, J. (2009). Easing the transition from student to doctor: how can medical schools help prepare their graduates for starting work? *Med Teach, 31*(5), 403-408. doi:10.1080/01421590802348127

Flexner, A. (1910). *Medical Education in the United States and Canada: A report to the Carnegie Foundation for the Advancement of Teaching*. New York City, NY: The Carnegie Foundation for the Advancement of Teaching.

Gaskell, N., Hinton, R., Page, T., Elvins, T., & Malin, A. (2016). Putting an end to Black Wednesday: improving patient safety by achieving comprehensive trust induction and mandatory training by day 1. *Clinical Medicine, 16*. doi:10.7861/clinmedicine.16-2-124.

Imperial College of London. (2009). *New study shows fresh thinking required on week junior doctors start*. Retrieved from https://www.imperial.ac.uk/news/

Jen, M. H., Bottle, A., Majeed, A., Bell, D., & Aylin, P. (2009). Early in-hospital mortality following trainee doctors' first day at work. *PLoS One, 4*(9), e7103. doi:10.1371/journal.pone.0007103

Miller, G. E. (1990). The Assessment of Clinical Skills/Competence/Performance. *Acad Med, 65*, s63 - s67. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2400509.

Morrison, E. W. (1993). Longitudinal Study of the Effects of Information Seeking on Newcomer Socialization. *Journal of Applied Psychology, 2*, 173-183. Retrieved from https://www.researchgate.net/.

Phillips, D. P., & Barker, G. E. (2010). A July spike in fatal medication errors: a possible effect of new medical residents. *J Gen Intern Med, 25*(8), 774-779. doi:10.1007/s11606-010-1356-3

Qureshi, Z., & Maxwell, S. (2012). Has bedside teaching had its day? *Adv Health Sci Educ Theory Pract, 17*(2), 301-304. doi:10.1007/s10459-011-9308-1

Sonnadara, R., McQueen, S., Mironova, P., Nousiainen, M., Ferguson, P., Alman, B., . . . Reznick, R. (2013). Reflections on current methods evaluating skills during joint replacement surgery. *The Bone & Joint Journal, 95-B*(11), 1445-1449. doi:10.1302/0301-620X.95B11

Sonnadara, R. R., Garbedian, S., Safir, O., Mui, C., Mironova, P., Nousiainen, M., . . . Reznick, R. (2013). Toronto orthopaedic boot camp III: examining the efficacy of student-regulated learning during an intensive, laboratory-based surgical skills course. *Surgery, 154*(1), 29-33. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23809482. doi:10.1016/j.surg.2013.05.003

Sonnadara, R. R., Garbedian, S., Safir, O., Nousiainen, M., Alman, B., Ferguson, P., . . . Reznick, R. (2012). Orthopaedic Boot Camp II: examining the retention rates of an intensive surgical skills course. *Surgery, 151*(6), 803-807. doi:10.1016/j.surg.2012.03.017

Sonnadara, R. R., Van Vliet, A., Safir, O., Alman, B., Ferguson, P., Kraemer, W., & Reznick, R. (2011). Orthopedic boot camp: examining the effectiveness of an intensive surgical skills course. *Surgery, 149*(6), 745-749. doi:10.1016/j.surg.2010.11.011

ten Cate, O. (2017). Competency Based Postgraduate Medical Education: Past Present and Future. *GMS Journal for Medical Education, 34*, 1-13. doi:10.3205/zma001146

Teunissen, P. W., & Westerman, M. (2011). Opportunity or threat: the ambiguity of the consequences of transitions in medical education. *Med Educ, 45*(1), 51-59. doi:10.1111/j.1365-2923.2010.03755.x

Van Maanen, J., & Schein, E. H. (1979). Toward A Theory of Organizational Socialisation. *Research in organizational behaviour, 1*, 209-264. Retrieved from https://dspace.mit.edu/bitstream/handle/1721.1/1934/?sequence=1.

Young, J. Q., Ranji, S. R., Wachter, R. M., Lee, C. M., Niehaus, B., & Auerbach, A. D. (2011). "July Effect": Impact of Academic Year-End Changeover on Patient Outcomes. *Annals of Internal Medicine, 155*, 309-315. Retrieved from htpp://www.annals.org.