Design and Validation of a Comprehensive Simulation-Enhanced Training Curriculum for a Complex Minimally Invasive Operation

by

Boris Zevin

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Institute of Medical Science
University of Toronto

© Copyright by Boris Zevin 2014
Design and Validation of a Comprehensive Simulation-Enhanced Training Curriculum for a Complex Minimally Invasive Operation

Boris Zevin
Doctor of Philosophy
Institute of Medical Science
University of Toronto
2014

Abstract

Laparoscopic bariatric procedures are complex minimally invasive operations with a potential for substantial morbidity and mortality along the early part of a surgeon’s learning curve. Simulation-enhanced training can improve a surgeon's technical and non-technical performance and lessen the learning curves in the operating room. Unfortunately, despite the convincing evidence supporting the use of simulation in surgical education, there is still a gap in translation of knowledge and technical skills from the research environment into clinically relevant training curricula. The objective of this thesis was to design and validate a comprehensive simulation-enhanced training curriculum that addressed cognitive knowledge, technical and non-technical skill in laparoscopic bariatric surgery. This objective was achieved using three experimental studies.

The first study employed a modified Delphi methodology and an international panel of experts in surgical and medical education to develop a consensus-based framework for design, validation and implementation of simulation-enhanced training curricula in surgery. The second study used a modified Delphi methodology and an international panel of experienced bariatric surgeons to
develop an objective scale for assessment of operative skill in laparoscopic gastric bypass procedure. This scale was feasible to use, had high inter-rater and test-retest reliability, as well as evidence of construct and concurrent validity.

The third study used the previously developed consensus-based framework to design a comprehensive simulation-enhanced training curriculum for laparoscopic bariatric surgery. A prospective, single-blinded randomized controlled trial was used to compare the effectiveness of this curriculum in comparison to conventional surgery training. Surgery residents who were trained in this curriculum demonstrated superior technical skills, superior non-technical skills and enhanced safety in the operating room.
Acknowledgments

I owe my sincerest gratitude to Dr. Teodor P. Grantcharov for allowing me to pursue this opportunity, for his continued support, guidance and mentorship. Of the many things that I have learned from you over the years, the two that I will cherish the most are “to never let a great opportunity pass by” and “to always remember which things are of real importance in life”.

I would like to extend my deepest appreciation to Dr. Najma Ahmed and Dr. Nicole Woods for their invaluable contributions as members of my thesis committee. Your helpful suggestions, keen attention to research methodology, constructive feedback and ongoing support were instrumental in raising the scientific quality of this work.

I wish to thank Dr. Rajesh Aggarwal for his invaluable contributions as a research collaborator and a mentor. I was always amazed at the number of great research ideas that were generated at each of our meetings.

I would also like to extend my sincerest appreciation to Drs. Esther Bonrath and Nicolas Dedy. Thank you for your ongoing support of my work, for always offering your help and most importantly for you friendship. I truly believe that we have many more years of productive research collaborations ahead.

I would like to thank the University of Toronto’s Department of Surgery and Division of General Surgery, Dr. Jim Rutka (Chair of the Department of Surgery) and Dr. Andy Smith (Chair of the Division of General Surgery) for their support and the opportunity to engage in the Surgeon Scientist Training Program.

I owe my deepest gratitude to my parents, Alexander and Natalia, for bringing me to the land of opportunities, for your continued encouragement, support, tremendous sacrifice and endless love.

I would like to acknowledge the following scholarship/fellowship programs: Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarship, Royal College of Physicians and Surgeons of Canada Fellowship for Studies in Medical Education, John L. Provan Fellowship in Surgical Education, and Ontario Graduate Scholarship.

The following agencies provided operating grants in support of this work: Physicians’ Services Incorporated Foundation Resident Research Grant, the Education Development Fund for Innovation in Education, and Johnson and Johnson Medical Companies, Canada.
Contributions

Dr. Boris Zevin (author) solely prepared this thesis. All aspects of this body of work, including the planning, execution, analysis, and writing of all original research and publications was performed in whole or in part by the author. The following contributions by other individuals are formally and inclusively acknowledged:

Dr. Teodor P. Grantcharov (Primary Supervisor and Thesis Committee Member) – mentorship; laboratory resources; guidance and assistance in planning, execution, and analysis of studies as well as manuscript/thesis preparation

Dr. Najma Ahmed (Thesis Committee Member) – mentorship; guidance in planning of studies and thesis preparation

Dr. Nicole Woods (Thesis Committee Member) – mentorship; guidance in planning of studies and thesis preparation

Dr. Rajesh Aggarwal – mentorship; guidance in planning and analysis of studies for Section 1.12 and Chapter 3

Dr. Esther Bonrath – assistance in planning, execution and analysis of studies for Chapters 3 and 4

Dr. Nicolas Dedy – assistance in planning, execution and analysis of studies for Chapters 3 and 4

Dr. Jeffrey Levy – assistance in execution of study for Chapter 2

Dr. Richard Satava – assistance in analysis of study for Chapter 2
Table of Contents

ACKNOWLEDGMENTS ........................................................................................................ IV

CONTRIBUTIONS ............................................................................................................. V

TABLE OF CONTENTS ...................................................................................................... VI

LIST OF TABLES ................................................................................................................. XII

LIST OF FIGURES ............................................................................................................. XIV

LIST OF ABBREVIATIONS ................................................................................................. XVI

CHAPTER 1 ......................................................................................................................... 1

1 LITERATURE REVIEW ................................................................................................. 2

1.1 Origins of Simulation in Surgery .................................................................................. 2

1.2 Surgical Simulation and Technical Skills Training ....................................................... 5
  1.2.1 Search Strategy ........................................................................................................ 5
  1.2.2 Bench-Top Trainers ............................................................................................... 5
  1.2.3 Virtual Reality Simulators ...................................................................................... 6
  1.2.4 Cadaveric Models .................................................................................................. 7
  1.2.5 Animal Models ....................................................................................................... 7
  1.2.6 Simulation Model Selection ................................................................................... 8

1.3 Educational Concepts in Surgical Education ............................................................... 12
  1.3.1 Search Strategy .................................................................................................... 12
  1.3.2 Deliberate Practice ............................................................................................... 12
  1.3.3 Practice Schedules .............................................................................................. 14

1.4 Objective Assessment Scales ...................................................................................... 16
  1.4.1 Search Strategy .................................................................................................... 16
  1.4.2 Checklists ............................................................................................................ 16
1.13.2 Surgical Simulation and Non-Feedback

1.5 Psychometric Properties of Assessment Scales

1.6 Feedback

1.7 Surgical Simulation and Non-Technical Skills Training

1.8 Theoretical Frameworks for Acquisition of Technical Skills

1.9 Transfer of Skill from the Laboratory to the Operating Room

1.10 Simulation-Enhanced Training and Patient Outcomes

1.11 Cost-Effectiveness of Simulation-Enhanced Training

1.12 Simulation-Based Training and Learning Curves in Laparoscopic Bariatric Surgery

1.13 The Delphi Method
CHAPTER 2 ....................................................................................................................... 69

2 A CONSENSUS-BASED FRAMEWORK FOR DESIGN, VALIDATION AND IMPLEMENTATION OF SIMULATION-BASED TRAINING CURRICULA IN SURGERY ......................................................... 70

2.1 Abstract ................................................................................................................... 70

2.2 Introduction .......................................................................................................... 71

2.3 Methods ............................................................................................................... 75
  2.3.1 Study Design .................................................................................................. 75
  2.3.2 Participant Selection ..................................................................................... 75
  2.3.3 Derivation of Survey Items ......................................................................... 78
  2.3.4 Survey Administration .................................................................................. 78
  2.3.5 Determination of Consensus ....................................................................... 78
  2.3.6 Creation of Final Framework ....................................................................... 78
  2.3.7 Data Analysis ................................................................................................ 82

2.4 Results .................................................................................................................. 83

2.5 Discussion ............................................................................................................ 90

2.6 Conclusion .......................................................................................................... 94

CHAPTER 3 .................................................................................................................... 96

3 DEVELOPMENT, FEASIBILITY, VALIDITY, AND RELIABILITY OF A SCALE FOR OBJECTIVE ASSESSMENT OF OPERATIVE PERFORMANCE IN LAPAROSCOPIC GASTRIC BYPASS SURGERY. 97

3.1 Abstract ............................................................................................................... 97

3.2 Introduction ........................................................................................................ 98
3.3 Methods............................................................................................................................................. 100
  3.3.1 Hierarchical Task Analysis............................................................................................................. 100
  3.3.2 Selection of Steps for Inclusion into the BOSATS Scale ............................................................... 101
  3.3.3 Determination of Feasibility of Use, Validity and Reliability of the BOSATS Scale ...................... 103

3.4 Results................................................................................................................................................ 106
  3.4.1 Hierarchical Task Analysis............................................................................................................. 106
  3.4.2 Selection of Steps for Inclusion into the BOSATS Scale ............................................................... 106
  3.4.3 Determination of Feasibility of Use, Validity and Reliability of the BOSATS Scale ...................... 121

3.5 Discussion........................................................................................................................................... 128

3.6 Conclusion......................................................................................................................................... 133

CHAPTER 4 ............................................................................................................................................... 134

4 DEVELOPMENT AND VALIDATION OF A COMPREHENSIVE SIMULATION-ENHANCED TRAINING CURRICULUM FOR A COMPLEX MINIMALLY INVASIVE OPERATION: A RANDOMIZED CONTROLLED TRIAL .................................................................................................................. 135

4.1 Abstract.............................................................................................................................................. 135

4.2 Introduction....................................................................................................................................... 137

4.3 Hypothesis......................................................................................................................................... 138

4.4 Methods........................................................................................................................................... 139
  4.4.1 Study Design & Dates ...................................................................................................................... 139
  4.4.2 Participants .................................................................................................................................... 140
  4.4.3 Inclusion and Exclusion Criteria ................................................................................................... 140
  4.4.4 Baseline Demographic and Clinical Data ...................................................................................... 141
  4.4.5 Baseline Assessment of Technical Skill ......................................................................................... 141
  4.4.6 Intervention .................................................................................................................................. 141
  4.4.7 Conventional Surgery Training (CST) .......................................................................................... 142
  4.4.8 Live Anesthetized Porcine Model .................................................................................................. 143
  4.4.9 Primary Outcome .......................................................................................................................... 143
  4.4.10 Secondary Outcomes .................................................................................................................. 143
  4.4.11 Sample Size ............................................................................................................................... 145
  4.4.12 Randomization ............................................................................................................................ 145
List of Tables

Table 1. Summary of retrospective case series defining the learning curve for LRYGB ........ 53

Table 2. Summary of prospective case series defining the learning curve for LRYGB .......... 53

Table 3. Surgery societies and organizations surveyed................................................. 76

Table 4. Composition of the Delphi international expert panel ..................................... 77

Table 5. Complete list of Delphi survey items.............................................................. 79

Table 6. An evidence-based framework for the design of a simulation-based surgical training curriculum ...................................................................................................................... 84

Table 7. Agreement among the participants on each component of the framework.......... 89

Table 8. List of all possible operative steps for the LRYGB ............................................ 107

Table 9. Composition of the international expert panel for the online Delphi questionnaire. . 112

Table 10. Agreement among the Delphi panel members for the 1st and 2nd round of online Delphi questionnaire ................................................................. 113

Table 11. The Bariatric Objective Structured Assessment of Technical Skill (BOSATS) scale 114

Table 12. Inter-rater reliability coefficients for the entire BOSATS scale and its individual components................................................................. 121

Table 13. Comparison of scores for different components of the LRYGB as assessed by the BOSATS scale using case-volume criteria for definition of experienced and novice surgeons................................................................................................................................. 122

Table 14. Comparison of scores for different components of the LRYGB as assessed by the BOSATS scale using OSATS GRS criteria for the definition of experienced (OSATS >= 28) and novice (OSATS < 28) surgeons......................................................... 124
Table 15. Spearman's rank correlation coefficients for comparisons between the BOSATS and the OSATS GRS scores for different components of the LRYGB. .......................... 126

Table 16. Demographic data and laparoscopic experience of the study participants............... 148

Table 17. Cognitive, technical and non-technical performance at baseline and post-training for the CST, SET and PGY 5 groups ......................................................................................................... 149

Table 18. Laparoscopic experience of the study participants during the course of the study.... 150

Table 19. Inter-observer agreement coefficients for skill categories on the NOTSS scale....... 157
List of Figures

Figure 1. Hypothetical model of attention................................................................. 3

Figure 2. Ranking of preferred simulation models by the year of training................... 9

Figure 3. Checklist for control of inferior vena cava hemorrhage .............................. 17

Figure 4. Objective Structured Assessment of Technical Skill Global Rating Scale........... 19

Figure 5. Modified Objective Structured Assessment of Technical Skill....................... 20

Figure 6. Global Operative Assessment of Laparoscopic Skills .................................. 21

Figure 7. Non-Technical Skills for Surgeons (NOTSS) scale ..................................... 22

Figure 8. Laparoscopic right colectomy procedure-specific assessment scale .............. 23

Figure 9. Oxford Non-technical Skills Scale ............................................................. 37

Figure 10. The additive effects of intrinsic and extraneous cognitive load .................... 40

Figure 11. Summary of studies reporting a change in duration of LRYGB operation following the completion of a predefined learning curve ................................................. 54

Figure 12. Framework for the design of a comprehensive ex-vivo training curriculum in laparoscopic bariatric surgery ................................................................. 61

Figure 13. Framework for systematic training and assessment of technical skills ............ 73

Figure 14. Suggested simulator curriculum for cognitive and manual skills training ......... 74

Figure 15. BOSATS scale scores for novice and experienced surgeon groups. Novice and experienced surgeons defined using case-volume criteria ................................. 123

Figure 16. BOSATS scale scores for novice and experienced surgeon groups. Novice and experienced surgeons defined using the Objective Structured Assessment of Technical Skills Global Rating Scale criteria ..................................................... 125
Figure 17. Spearman's rank correlation for the BOSATS and the OSATS GRS scores for the creation of hand sewn GJ anastomosis ................................................................. 127

Figure 18. Study Flow Diagram ........................................................................................................................................ 139

Figure 19. CONSORT Diagram .......................................................................................................................................... 140

Figure 20. The Bariatric Objective Structured Assessment of Technical Skill (BOSATS) scale for the creation of laparoscopic jejunojejunostomy ............................................. 144

Figure 21. Dot plot of the technical skill scores at baseline assessment for the CST and the SET group .................................................................................................................. 151

Figure 22. Dot plot of the technical skill scores for the CST and the SET group for the 1st jejunojejunostomy in a live anesthetized porcine model .................................................. 152

Figure 23. Dot plot of the technical skill scores for the CST and the SET group for the jejunojejunostomy in the operating room ........................................................................... 154

Figure 24. Dot plot of the technical skill scores for the CST group at baseline, on the 1st jejunojejunostomy in the live anesthetized porcine model and the jejunojejunostomy in the operating room .................................................................................. 155

Figure 25. Dot plot of the technical skill scores for the SET group at baseline, on the 1st jejunojejunostomy in the live anesthetized porcine model and the jejunojejunostomy in the operating room .................................................................................. 156

Figure 26. Dot plot of non-technical skill scores for post-intervention assessment of the CST and the SET group .......................................................................................................... 158

Figure 27. Dot plot of the technical skill scores for jejunojejunostomy in the operating room for the CST, the SET, and the PGY 5 group ........................................................................ 160
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACGME</td>
<td>Accreditation Council for Graduate Medical Education</td>
</tr>
<tr>
<td>ACS</td>
<td>American College of Surgeons</td>
</tr>
<tr>
<td>ADEPT</td>
<td>Advanced Dundee Psychomotor Tester</td>
</tr>
<tr>
<td>APDS</td>
<td>Association for Program Directors in Surgery</td>
</tr>
<tr>
<td>ASA</td>
<td>American Society of Anesthesiology</td>
</tr>
<tr>
<td>ASMBS</td>
<td>American Society for Metabolic and Bariatric Surgery</td>
</tr>
<tr>
<td>ASSET</td>
<td>Alliance for Surgical Simulation Education and Training</td>
</tr>
<tr>
<td>ATLAS</td>
<td>Advanced Training in Laparoscopic Abdominal Surgery</td>
</tr>
<tr>
<td>BMI</td>
<td>Body-Mass Index</td>
</tr>
<tr>
<td>BOSATS</td>
<td>Bariatric Objective Structured Assessment of Technical Skills</td>
</tr>
<tr>
<td>CAD</td>
<td>Canadian Dollar</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CST</td>
<td>Conventional Surgery Training</td>
</tr>
<tr>
<td>ES</td>
<td>Effect Size</td>
</tr>
<tr>
<td>FLS</td>
<td>Fundamentals of Laparoscopic Surgery</td>
</tr>
<tr>
<td>GJ</td>
<td>Gastrojejunostomy</td>
</tr>
<tr>
<td>GOALS</td>
<td>Global Operative Assessment of Laparoscopic Skills</td>
</tr>
<tr>
<td>GRS</td>
<td>Global Rating Scale</td>
</tr>
<tr>
<td>HTA</td>
<td>Hierarchical Task Analysis</td>
</tr>
<tr>
<td>HUESAD</td>
<td>Hiroshima University Endoscopic Surgical Assessment Device</td>
</tr>
<tr>
<td>ICSAD</td>
<td>Imperial College Surgical Assessment Device</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass Correlation Coefficient</td>
</tr>
<tr>
<td>JJ</td>
<td>Jejunojejunostomy</td>
</tr>
<tr>
<td>LRYGB</td>
<td>Laparoscopic Roux-en-Y Gastric Bypass</td>
</tr>
<tr>
<td>MeSH</td>
<td>Medical Subject Headings</td>
</tr>
<tr>
<td>MISTELS</td>
<td>McGill Inanimate System for Training and Evaluation of Laparoscopic Skills</td>
</tr>
<tr>
<td>MIST-VR</td>
<td>Minimally Invasive Surgical Trainer Virtual Reality</td>
</tr>
<tr>
<td>NOTSS</td>
<td>Non-Technical Skills for Surgeons</td>
</tr>
<tr>
<td>NOTACHS</td>
<td>NO-TECHnical Skills</td>
</tr>
<tr>
<td>OR</td>
<td>Operating Room</td>
</tr>
<tr>
<td>OSATS</td>
<td>Objective Structured Assessment of Technical Skills</td>
</tr>
<tr>
<td>OSCE</td>
<td>Objective Structured Clinical Encounter</td>
</tr>
<tr>
<td>OTAS</td>
<td>Observational Teamwork Assessment for Surgery</td>
</tr>
<tr>
<td>PGY</td>
<td>Postgraduate Year</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized Controlled Trial</td>
</tr>
<tr>
<td>SET</td>
<td>Simulation-Enhanced Training</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>TeamSTEPPS</td>
<td>Team Strategies and Tools to Enhance Performance and Patient Safety</td>
</tr>
<tr>
<td>TER</td>
<td>Transfer Effectiveness Ratio</td>
</tr>
<tr>
<td>THC</td>
<td>Thiel Human Cadavers</td>
</tr>
<tr>
<td>USD</td>
<td>United States Dollar</td>
</tr>
<tr>
<td>VR</td>
<td>Virtual Reality</td>
</tr>
</tbody>
</table>
Chapter 1

Literature Review
1 Literature Review

1.1 Origins of Simulation in Surgery

The apprenticeship model of surgical training was introduced to North America in early 20th century. This introduction is largely credited to Professor William Stewart Halstead during his time at Johns Hopkins Hospital, Baltimore, Maryland, United States (Rankin, 2006). In this model, surgery trainees were supervised by an attending surgeon and were offered a graduated level of responsibility as their training progressed. Most of the practice and training occurred on real patients in the operating room. This model of training has indeed produced many of the outstanding surgeons of the 20th century; however, this approach to surgery training may no longer be suitable in the 21st century in view of the minimally invasive approaches to conventional surgical procedures, reductions in resident work hours, decreases in the available operating room time, increases in the risks of malpractice litigation, increases in patient complexity and ethical requirements to protect patients from harm (Palter & Grantcharov, 2010). Simulation-based training has been proposed as a useful adjunct to conventional surgery training to deal with some of these challenges that are facing surgeons of the 21st century (Torkington, Smith, Rees, & Darzi, 2000).

Simulation-based training and practice in a simulation laboratory offers a number of advantages to a surgery trainee. It allows a trainee to practice a particular surgical task without any of the time pressures and inherent risks of training in the operating room. A trainee is allowed to make mistakes, correct them and try different approaches without any risk to the patient. He or she can become accustomed to various surgical instruments and can build the necessary dexterity for active participation in the operating room. An experienced attending surgeon can often notice that a novice surgery trainee is often not able to acquire new information or carry on a conversation in the operating room, while he or she is performing a basic surgical task. The explanation for this phenomenon centers on the availability cognitive resources at the time of performing a technical task (Figure 1). The cognitive resources of a novice surgery trainee are completely depleted by the technical task at hand, thus, he or she is not able to perform and engage in any other tasks in the operating room.
Training in a simulation laboratory offers a novice trainee an opportunity to mature into a “pre-trained novice” (Gallagher et al., 2005). A pre-trained novice is an individual who has been trained using simulation to the point where many of the psychomotor skills and spatial judgments have been automated and occupy fewer cognitive resources, allowing a novice trainee to focus more on higher level learning in the operating room (Figure 1). Palter et al. confirmed this hypothesis in a single-blinded randomized controlled trial of 18 novice surgery trainees randomized to either the simulation-based training group or the control group (Palter, Grantcharov, Harvey, & Macrae, 2011). Following randomization, both groups had an introductory session on the basics of fascial closure. The simulation-based training group then practiced fascial closure to proficiency on an inanimate model, whereas the control group did not. At the completion of training both groups performed a fascial closure in the operating room, while listening to a verbal script that contained relevant clinical information. Technical skill in the operating room was assessed using the Objective Structured Assessment of Technical Skill (OSATS) scale, whereas the acquired knowledge from the script was assessed using a multiple-choice test. Participants in the simulation-based training group showed significantly greater technical skill in the operating room and knowledge on the multiple-choice test.
Training in a simulation laboratory provides trainees with the opportunity to engage in deliberate practice. According to K.A. Ericsson, deliberate practice is the key to expertise in professional sports, music, chess and medicine (Ericsson, 2008). Deliberate practice requires (a) a task with a well-defined goal, (b) motivation to improve, (c) provision of feedback, and (d) ample opportunities for repetition and gradual refinements in performance (Ericsson, 2008). Training on surgical simulators is well suited for deliberate practice as well defined goals can be set, feedback can be provided and multiple repetitions of a task can be performed at increasing levels of difficulty.
1.2 Surgical Simulation and Technical Skills Training

1.2.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


The first surgical simulators began appearing in the early 1990’s. Professor Richard M. Satava was one of the first to describe a virtual reality simulator for surgical skills training back in 1993 (Satava, 1993). Over the past 2 decades many different simulators were developed for use in technical and non-technical skills training. These include bench-top trainers, virtual reality simulators, models based on cadaveric tissues and live anesthetized animal models.

1.2.2 Bench-Top Trainers

Bench-top trainers are surgical simulators, which can be used for acquisition of open and laparoscopic (minimally invasive) surgical skills. They are frequently made from synthetic materials and allow for practice of surgical tasks such as suturing, tissue handling, Foley catheter insertion, chest tube placement and others. Bench-top trainers were successfully used to create an examination for objective structured assessment of technical skills (OSATS) (J. A. Martin et al., 1997), which included the following open surgical skills: excision of skin lesion, hand-sewn bowel anastomosis, stapled bowel anastomosis, insertion of T-tube, abdominal wall closure and control of inferior vena cava hemorrhage.

Bench-top trainers have also been used in the training and objective assessment of laparoscopic surgical skills. Such trainers - often referred to as video-box trainers - are made up of a box with multiple holes for placement of a laparoscopic camera and laparoscopic instruments. Synthetic and cadaveric tissues can be placed inside this box and various minimally invasive surgical skills (grasping, cutting, suturing, etc.) can be practiced. The technical component of the Society for
American Gastrointestinal and Endoscopic Surgeons (SAGES) and the American College of Surgeons (ACS) Fundamentals of Laparoscopic Surgery (FLS) curriculum relies primarily on a video-box trainer for training and assessment of the following basic laparoscopic skills: dexterity (peg transfer), cutting (pattern cutting), placement of a ligating loop and suturing (intracorporeal and extracorporeal knot placement) (Swanstrom, Fried, Hoffman, & Soper, 2006). Complex laparoscopic skills can also be learned and assessed using a video-box trainer with synthetic and cadaveric tissues. Such models are already available for a laparoscopic Nissen fundoplication (S. M. Botden, L. Christie, R. Goossens, & J. J. Jakimowicz, 2010), a laparoscopic inguinal hernia repair (Slater et al., 2001) and a laparoscopic jejunojejunostomy (R. Aggarwal, Boza, et al., 2007).

Bench-top trainers have a number of advantages and disadvantages in comparison to other training modalities (Munz, Kumar, Moorthy, Bann, & Darzi, 2004). Bench-top trainer advantages include low cost, versatility (can train on synthetic or cadaveric tissues), ability to use real instruments, and standardization of training. Bench-top trainers are also well suited for use in assessment of technical skills. The disadvantages of bench-top trainers include inability to simulate a complete operation, difficulties in adjusting the level of difficulty for deliberate practice, requirement for ongoing maintenance of equipment, and the requirement for presence of an experienced instructor for demonstration of skills and provision of feedback.

1.2.3 Virtual Reality Simulators

Virtual reality (VR) simulators use a computer interphase to generate three-dimensional images of anatomy, objects and surgical instruments in a virtual space. The first VR simulators began to appear in the early 1990’s (Satava, 1993). Over the past two decades, the number of available VR surgical simulators has grown exponentially. Currently, there are simulators for basic laparoscopic tasks, advanced laparoscopic tasks, and procedural training in general surgery, orthopedics, otolaryngology, urology, vascular surgery and obstetrics and gynecology. Newer VR models feature haptic feedback, which is a feature that provides a user with the illusion of coming into physical contact with the model. Consensus guidelines for validation of VR simulators have been developed (Carter et al., 2006) and a substantial amount of work has been done to demonstrate construct validity of various VR simulators (Duffy et al., 2005; van Dongen, Tournoij, van der Zee, Schijven, & Broeders, 2007; Woodrum et al., 2006).
The advantages of VR simulators include the ability to simulate surgical complications and entire surgical procedures, provision of automatic objective feedback without an experienced instructor being present, minimal requirement for maintenance and an ability to change the level of difficulty for deliberate practice (Munz et al., 2004). The disadvantages include high up-front set-up costs, lack of realistic graphics and limited force feedback (Munz et al., 2004).

1.2.4 Cadaveric Models

Human cadavers have been used in teaching of surgical skills for more than a century. Fortunately, the demand for human cadavers as a teaching tool has been decreasing with the development of bench-top trainers and VR simulators. The advantages of cadaver models include high fidelity, realistic anatomy and the ability to simulate entire operations. The disadvantages include high cost, limited availability, risk of transmissible infections, lack of bleeding, inability to simulate complications, and poor tissue compliance (Munz et al., 2004; Palter & Grantcharov, 2010). The disadvantages in poor tissue compliance can be overcome with the use of Thiel human cadavers (Giger, Fresard, Hafliger, Bergmann, & Krahenbuhl, 2008), albeit such preparations make these models even more expensive. At the present time, training on human cadavers may be best reserved for advanced surgical skills and entire surgical procedures (Giger et al., 2008).

1.2.5 Animal Models

Cadaveric animal tissue models and live anesthetized animals have been used in surgical skills training. Cadaveric tissue models have been used in laparoscopic cholecystectomy, endoscopic retrograde cholangiopancreatography and bronchoscopy training (Palter & Grantcharov, 2010). The advantages of these models include the ease of procurement, good tissue handling properties when fresh and a relatively low price (Munz et al., 2004). The disadvantages include anatomic differences from human anatomy, risk of infection, inherent costs of maintaining appropriate facilities with trained personnel, difficulties with procurement, as well as ethical considerations in using cadaver models for surgery training.

Live anesthetized animal models are still being used in some countries for surgical skills training. Advantages of live animal models include high fidelity and realism of tissues, excellent tissue compliance, opportunities to simulate intra-operative complications and practice operations in their entirety, as well as the ability to induce bleeding for vascular procedures (Munz et al., 2004). The disadvantages include inability to use live anesthetized animals for
surgery training in some countries such as the United Kingdom, high acquisition and maintenance costs, ethical concerns about practicing on live animals, and differences from human anatomy (Munz et al., 2004).

1.2.6 Simulation Model Selection

Despite the wide array of available models for surgical simulation, it is still unclear which model should be used in which context. Does every task require a high fidelity simulator or can certain tasks be taught using low fidelity and low cost simulators? Are VR simulators more effective than video-box trainers for teaching basic laparoscopic skills? Which simulators are most cost-effective and in which context? What are the preferences of trainees and do they differ by year of training? The answers to some of these questions are already available; others will require further scientific inquiry.

In a recent study, we conducted an online questionnaire of 67 post-graduate year (PGY) 1 to 5 general surgery residents at Yale University School of Medicine and University of Toronto inquiring about resident’s perceptions, training experiences, and preferences regarding laparoscopic simulation training (Shetty et al., 2013). As a group (PGY 1 to 5), residents ranked live animal models as the most preferred training modality, followed by cadaver animal tissues, video-box trainers and VR simulators. Interestingly, rankings varied by year of training (Figure 2).
Figure 2. Ranking of preferred simulation models by the year of training.

Junior level trainees (PGY 1) preferred video-box trainers and VR simulators. Intermediate level trainees (PGY 2-3) preferred live animal models and video-box trainers with cadaver tissues. Senior level trainees (PGY 4-5) preferred training on live animal models. These results suggest that junior level trainees are interested in practicing basic surgical skills on low fidelity video-box trainers and VR simulators, whereas senior level trainees are more interested in practicing operations in their entirety with potential intraoperative complications, thus preferring to use live anesthetized animal models.

Munz and colleagues investigated whether VR simulators were more effective than video-box trainers at teaching basic laparoscopic skills using a prospective randomized controlled trial of 24 novices (medical students) randomly allocated to one of three groups: box trainer, VR simulator or control (Munz et al., 2004). Participants were trained in the box trainer or the VR simulator in basic laparoscopic tasks over 3 weekly sessions of 30 minutes each. Technical skills were objectively assessed on a box trainer using the Imperial College Surgical Assessment Device (ICSAD). Outcome metrics included the total number of hand movements, total distance traveled for each hand, total time taken to perform the task and the economy of movement.
Statistically significant improvements in the economy of hand movement, path length and number of errors were seen in the box trainer and VR trainer groups. There was no difference in skill improvement between these two groups. The authors concluded that box trainers and VR simulators were equally effective at improving technical skills in novices.

In a prospective single-blinded randomized controlled trial of 24 non-novice surgery trainees (PGY 2 and above), Orzech and colleagues compared proficiency-based training in laparoscopic intracorporeal suturing on a VR simulator versus box trainer (Orzech, Palter, Reznick, Aggarwal, & Grantcharov, 2012). Technical skill was assessed on a synthetic Nissen fundoplication model and intra-operatively using a validated procedure-specific scale and a global rating scale. Results demonstrated no significant differences between box trainer and VR simulator groups in time to complete the suture, as well as procedure-specific scores and global rating scores. Based on these results, VR simulator and box trainer were equally effective for teaching laparoscopic suturing. Interestingly, this study also used the Transfer Effectiveness Ratio (TER) to examine the effectiveness of each model. TER was defined using the following equation: 

\[
\text{TER} = \frac{Y_0 - Y_x}{X};
\]

where \( Y_0 \) is the median time required by the control (untrained) group to reach a predefined performance criterion, and \( Y_x \) represents the corresponding measure for the intervention group after having received a median of \( X \) amount of training time on a simulator. The TER for VR simulator and box trainer was 2.31 and 1.13, respectively; suggesting that training on a VR simulator was more efficient. This study also reported on some crude calculations of cost-effectiveness for VR and box trainer in comparison to the untrained group. The annual cost of training 5 residents on the box trainer was reported as $11,975.00, on the VR simulator as $77,500.00, and in conventional residency as $17,380.00. The authors concluded, “VR training was the more efficient training modality, whereas box training was the more cost-effective option”.

A recent systematic review and meta-analysis of clinical trials up to May 11, 2011 compared simulation-based training in laparoscopic surgery to no intervention, to a non-simulation intervention and to a different simulation-based intervention (Zendejas, Brydges, Hamstra, & Cook, 2013). The outcomes of interest were: participants’ satisfaction, skill (in a test setting), and behavior when caring for patients. The outcome of skill was subdivided into time (to perform the task), process (e.g. performance rating), and product (e.g. knot strength). Twenty-one studies (17 randomized controlled trials) compared VR simulators to box trainers. The pooled effect size favored the box trainer for the outcomes of participants’ satisfaction (effect
size 0.61) and time (effect size 0.33). Interestingly, there were no significant differences between VR simulator and box trainer for all other outcomes. Two studies compared box trainers and animal models; two other studies compared box trainers and human cadaver models. In all four of these studies, the box trainer was favored for the outcome of skill (effect size 0.79). Based on the results of this systematic review, box trainers seem to be at least equivalent, if not superior, to other simulated training modalities when evaluated for the variables of participants’ satisfaction, skill, and behavior when caring for patients. This study did not, however, address the cost-effectiveness of different simulation modalities.
1.3 Educational Concepts in Surgical Education

1.3.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


All simulation-based training sessions must be underpinned by the educational concepts of deliberate practice and a distributed practice schedule regardless of the type of simulator in use.

1.3.2 Deliberate Practice

Professor K. Anders Ericsson has stated that expertise should not be ascribed to individuals with greatest experience, greatest accumulated knowledge or greatest number of peer nominations; rather, it should be ascribed to those individuals who demonstrate “reproducibly superior performance” (Ericsson, 2008). Applying this statement to the domain of surgery suggests that expert surgeons are those individuals that have reproducibly superior patient outcomes, rather than those surgeons that have performed the greatest number of operations. Deliberate practice is required and has been uniformly associated with improved performance in sports, music and medicine (Ericsson, 2004). According to Ericsson, four conditions are required for deliberate practice: (1) provision of a task with a well-defined goal; (2) motivation of an individual to improve; (3) provision of feedback; and (4) provision of ample opportunities for repetition and gradual refinement of performance.

To understand the mechanism by which deliberate practice leads to superior performance, we must discuss Dreyfus and Dreyfus’s five stages of skill acquisition and the concept of “automaticity” in technical skills. Dreyfus and Dreyfus’s five stages of skill acquisition are: novice, advanced beginner, competent, proficient and expert (Dreyfus, Dreyfus, & Athanasiou,
For a novice learner to learn a technical skill, the instructional process must begin by decomposing the task into context-free features and providing the rules for determining actions on the basis of these features. A useful illustrative example of this instructional process can be made for a medical student who is learning how to suture. The task of suturing can be decomposed into context-free rules of loading the needle, pronation of the wrist, entering tissue at 90 degrees, supination of the wrist and reloading the needle. These context-free rules are then taught to a medical student at the novice stage of skill acquisition. As the medical student gains increasing experience with real situations, he or she will learn to recognize aspects of these situations and will modify prior context-free rules according to these newly acquired aspects. With increasing situational experience, a novice medical student on a surgery rotation will build on the context-free rules of suturing by learning which type of suture and what type of stitch is useful in which situation. With the recognition of situation-specific aspects, a “novice” trainee reaches an “advanced beginner” stage of skill acquisition.

As an individual’s situational experience increases, the number of situational aspects and cues to be taken into account becomes overwhelming. To adapt to this information overload, an “advanced beginner” begins to create a hierarchical view of decision-making and transitions to a “competence” level. He or she first chooses a perspective or a plan to organize a given situation and then examines only a small set of aspects (learned from prior experiences) that are relevant to that plan. The decision-making in this stage remains analytical; however, this hierarchical view allows an individual to simplify and improve his or her performance. An “advanced beginner” surgery resident who has closed a number of laparotomy incisions will begin to recognize which wounds are prone to dehiscence and may decide to modify his or her technique from a continuous running suture to simple interrupted sutures in certain situations. As this occurs, the resident transitions from an “advanced beginner” to “competence” stage of skill acquisition.

A “competent” individual who becomes “proficient” stops looking for principles to guide his or her actions and rather uses holistic experiences from the “competent” stage to guide his or her actions. The understanding of what is going on in a particular situation becomes effortless; however, the decision of what to do in that situation remains analytical. A “proficient” individual progresses to an “expert” stage by having enough experience with a variety of different classes of situations, which have been decomposed into subclasses associated with the same decision and action. An “expert” does not need to think of how to close an abdomen; he does it
intuitively and automatically. “Automaticity” in technical skill is reached at this stage. Responses become rapid and intuitive, and additional experience does not result in a greater level of performance (Ericsson, 2008). Thus, additional improvement in technical skill can only result from deliberate practice, which allows an individual to avoid the arrested development associated with automaticity by generating more challenging performance goals that exceed current level of performance.

Training on surgical simulators provides an opportunity for deliberate practice. Features of simulation training that contribute to deliberate practice include: (a) presence of well-defined learning objectives or tasks; (b) ability to select various levels of difficulty for each task; (c) opportunity for focused and repetitive practice; (d) presence of rigorous and precise educational measurements; and (e) availability of feedback (McGaghie, 2008). Task difficulty can be changed on a VR simulator; position of the monitor and/or the angle of laparoscopic camera will increase the level of difficulty for tasks in a box trainer (Haveran et al., 2007).

1.3.3 Practice Schedules

The discussion of massed (all at one time) versus distributed (spaced across a number of sessions) practice schedules for learning of motor skills has been present in the kinesiology literature since the 1960’s. A meta-analysis of 63 studies in the domain of kinesiology suggested that distributed practice is superior to massed practice with a mean weighed effect size of 0.46 (Donovan & Radosevich, 1999). Interestingly, when the authors stratified the results of the meta-analysis based on the type of motor task (from a low mental requirement, low physical requirement and low overall complexity to a high mental requirement, high physical requirement and high overall complexity) and the time interval between practice sessions (< 1 minute, 1-10 minutes, 10 minutes – 1 hour, > 1 hour), the effect sizes for distributed practice schedule diminished with increased task complexity and increasing time interval between practice sessions. Moulton and colleagues replicated these findings for the acquisition of microvascular surgical skills (Moulton et al., 2006). In their study, 38 junior surgery trainees were randomized to either a massed (4 training sessions in 1 day) or a distributed (1 session per week for 4 weeks) practice schedule. Surgical skills were taught with the aid of instructional videos and synthetic bench-top models. Each study participant underwent baseline testing, post-intervention testing, retention testing (1-month after completion of the training) and transfer testing (performance of a microvascular anastomosis on a live anesthetized rat model). Expert-based metrics (global rating scale, procedure-specific checklist, competency rating, final product analysis and clinically
relevant measures, such as bleeding, patency, narrowing and completion rate) and computer-based metrics (time and number of movements) were used for assessment of technical skill. Significant between-group differences in favor of the distributed practice schedule group were noted for retention testing (global rating scale and competency rating) and transfer testing (global rating, checklist, final product analysis and competency score). Moulton and colleagues concluded that surgical skills acquired under distributed practice conditions are more robust and show superior transferability to a lifelike model than surgical skill acquired under massed conditions.

So, why is a distributed practice schedule more effective than a massed practice schedule? The answer to this question centers on the concepts of consolidation, interference and self-rehearsal. Consolidation is defined as a process by which long-term memory becomes more stable with the passage of time (Krakauer & Shadmehr, 2006). Interference is defined as a process by which training of a new task leads to forgetting of a previously learned task (Krakauer & Shadmehr, 2006). Consolidation of one task can be disrupted if another task is learned right after (Brashers-Krug, Shadmehr, & Bizzi, 1996). In the case of a distributed practice schedule, the presence of an interruption in the training schedule between learning of the first task and the second task has been shown to decrease interference and lead to greater consolidation of the first task (Krakauer & Shadmehr, 2006). Interestingly, rest between training sessions seems to lead to additional performance gains via some form of “self-rehearsal” (Walker, Brakefield, Hobson, & Stickgold, 2003). A distributed practice schedule, therefore, diminishes interference, improves consolidation and allows for “self-rehearsal” or “mental practice” of a task during the rest intervals. This process of “self-rehearsal” of a task from memory prior to each practice session is likely partially responsible for deeper encoding and better retention of learned motor skills.
1.4 Objective Assessment Scales

1.4.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


Objective assessment of technical and non-technical skills in surgery can (1) aid learning by provision of constructive feedback, (2) determine the level that a trainee has achieved, (3) check whether a trainee has made progress, (4) ensure patient safety before a trainee performs an unsupervised procedure, and (5) certify completion of training (Beard, 2007). A number of different objective assessment scales are currently available for evaluation of technical and non-technical skills in surgery. These scales can be grouped into the categories of checklists, global rating scales, procedure-specific scales, motion analysis and virtual reality simulators. A systematic review of methods for objective assessment of surgical skills by Van Hove and colleagues provided an excellent overview of different assessment scales, their validity and reliability (van Hove, Tuijthof, Verdaasdonk, Stassen, & Dankelman, 2010). Herein, I will highlight the main advantages and disadvantages of each type of scale.

1.4.2 Checklists

Different types of checklists are currently available for objective assessment of technical skills in a simulated environment (J. A. Martin et al., 1997) and in the operating room ("ACS/APDS Surgical Skills Curriculum for Residents,"). The American College of Surgeons (ACS) and the Association for Program Directors in Surgery (APDS) has created checklists for some of the more common general surgery operations including appendectomy, cholecystectomy, Nissen fundoplication, ventral hernia repair and others.
The usual process to create a procedure-specific checklist involves a panel of experts in the field that agrees on the list of surgical steps that must be completed in order to perform a specific procedure. These selected steps are then used as items on the checklist, each with a categorical response of "completed" or "not completed". Thus, a checklist is simply completed by observing a surgical procedure and checking off the items that were completed. This rigid categorical structure of the checklist is in part responsible for its advantages and disadvantages for the assessment of surgical skill (Regehr, MacRae, Reznick, & Szalay, 1998).

The main advantages of checklists include objectivity, unambiguous expectations and the opportunity to provide immediate and relevant feedback (Regehr et al., 1998). Using the example of the checklist in Figure 3, a trainee can be informed that he did not load the needle appropriately during suturing, and this feedback can be used to modify his future behavior. Checklists also turn examiners into observers of behavior rather than interpreters of behavior, thereby making the assessment more objective and less subjective.

### CHECKLIST

<table>
<thead>
<tr>
<th>ITEM</th>
<th>Not Done/ Done Inorrectly</th>
<th>Done Correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONTROL OF HEMORRHAGE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Applies pressure to stop bleeding first</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2. Asks assistant to suction field</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3. Inspects injury by carefully releasing the IVC</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4. Ensures all equipment needed for repair is at hand before starting</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5. Control of bleeding point (use deBakey forceps / Satinsky clamp or prox/distal pressure)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>REPAIR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Select appropriate suture (4.0/5.0/6.0 polypropylene)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7. Select appropriate needle driver (vascular)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8. Select appropriate forceps (de Bakey)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9. Needle loaded 1/2-2/3 from tip 90% of time</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Checklists also have several disadvantages. First, every unique operative procedure requires a unique procedure-specific checklist. In other words, checklists are not generalizable from one procedure to another. Second, checklists tend to reward thoroughness - ability to complete every item on the checklist - rather than completeness. They do not recognize alternative approaches to an operative procedure and do not allow an experienced rater to make a “holistic” judgment of an individual performance (Regehr et al., 1998). The rigid structure of the checklist may allow a novice trainee who completes all of the steps of a particular procedure to achieve a greater score than an experienced surgeon who does not complete each step of the procedure listed on the checklist yet achieves a more complete final product (Norman, Tugwell, Feightner, Muzzin, & Jacoby, 1985). It is well known that experienced surgeons and clinicians take shortcuts to get to a complete final product; however, this approach would not be reflected in the score on the checklist.

1.4.3 Global Rating Scales

Global rating scales (GRS) offer another method for objective assessment of technical and non-technical skills in surgery. Examples of GRS for assessment of technical skills include Objective Structured Assessment of Technical Skill Global Rating Scale (OSATS GRS) in Figure 4 (J. A. Martin et al., 1997), modified OSATS GRS in Figure 5 (Grantcharov et al., 2004a), and the Global Operative Assessment of Laparoscopic Skills (GOALS) in Figure 6 (Vassiliou et al., 2005). Examples of GRS for assessment of non-technical skills of include the Non-Technical Skills for Surgeons (NOTSS) scale in Figure 7 (Yule et al., 2008; Yule, Flin, Paterson-Brown, Maran, & Rowley, 2006), the NON-TECHnical Skills (NOTECHS) scale (Sevdalis et al., 2008) and the Observational Teamwork Assessment for Surgery (OTAS) scale (Sevdalis et al., 2009).

The main advantage of a global rating scale over a checklist is the ability to evaluate different procedures using the same GRS, because GRS are designed to assess general operative skills such as tissue handling, knowledge of instruments and others. These general operative skills are transferable from one procedure to the next. Checklists, on the other hand, evaluate whether specific operative steps were completed correctly, making them not transferable from one procedure to the next. As an example, OSATS GRS has been used for the assessment of technical skill during chest tube insertion, excision of a skin lesion and various procedures in the operating room (J. A. Martin et al., 1997; Niitsu et al., 2012).
The main disadvantages of global rating scales include the requirement to train the assessors in the use of the scale (Feldman, Lazzara, Vanderbilt, & DiazGranados, 2012), introduction of some subjectivity during the assessment, and the difficulty with using the scores for formative (constructive) feedback. If one examines the OSATS GRS in Figure 4, it is clear that rater training would be required to judge what is “competent use of instruments” and what is “good use of assistants”. An individual who is not familiar with the discipline of surgery would require extensive training in order to score such a GRS correctly. A judgment of what is “competent use of instruments” introduces a subjective component into this objective assessment scale. The question does not simply ask if a step was completed or not, as is the case with a checklist, rather it requires the assessor to make a judgment, which is susceptible to bias. Lastly, a score on a global rating scale would be a less useful in terms of formative feedback to a trainee. For example, a trainee that achieved a low score on the OSATS GRS will not know which part of the operation he or she should focus on for improvement of technical skills.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for tissue</td>
<td>Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments</td>
<td>Careful handling of tissue but occasionally caused inadvertent damage</td>
<td>Consistently handled tissues appropriately with minimal damage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time and motion</td>
<td>Many unnecessary moves</td>
<td>Efficient time/motion but some unnecessary moves</td>
<td>Economy of movement and maximum efficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrument handling</td>
<td>Repeatedly makes tentative or awkward moves with instruments</td>
<td>Competent use of instruments although occasionally appeared stiff or awkward</td>
<td>Fluid moves with instruments and no awkwardness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of instruments</td>
<td>Frequently asked for the wrong instrument or used an inappropriate instrument</td>
<td>Knew the names of most instruments and used appropriate instrument or the task</td>
<td>Obviously familiar with the instruments required and their names</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of assistants</td>
<td>Consistently placed assistants poorly or failed to use assistants</td>
<td>Good use of assistants most of the time</td>
<td>Strategically used assistant to get the best advantage at all times</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flow of operation and forward planning</td>
<td>Frequently stopped operating or needed to discuss next move</td>
<td>Demonstrated ability for forward planning with steady progression of operative procedure</td>
<td>Obviously planned course of operation with effortless flow from one move to the next</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of specific procedure</td>
<td>Deficient knowledge. Needed specific instruction at most operative steps</td>
<td>Knew all important aspects of the operation</td>
<td>Demonstrated familiarity with all aspects of the operation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Modified Objective Structured Assessment of Technical Skill (obtained from Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg.* Feb 2004;91(2):146-150)
Global rating scale component of the intraoperative assessment tool*

Depth perception
1. Constantly overshoots target, wide swings, slow to correct
2.
3. Some overshooting or missing of target, but quick to correct
4.
5. Accurately directs instruments in the correct plane to target

Bimanual dexterity
1. Uses only one hand, ignores nondominant hand, poor coordination between hands
2.
3. Uses both hands, but does not optimize interaction between hands
4.
5. Expertly uses both hands in a complimentary manner to provide optimal exposure

Efficiency
1. Uncertain, inefficient efforts; many tentative movements; constantly changing focus or persisting without progress
2.
3. Slow, but planned movements are reasonably organized
4.
5. Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach

Tissue handling
1. Rough movements, tears tissue, injures adjacent structures, poor grasper control, grasper frequently slips
2.
3. Handles tissues reasonably well, minor trauma to adjacent tissue (i.e., occasional unnecessary bleeding or slipping of the grasper)
4.
5. Handles tissues well, applies appropriate traction, negligible injury to adjacent structures

Autonomy
1. Unable to complete entire task, even with verbal guidance
2.
3. Able to complete task safely with moderate guidance
4.
5. Able to complete task independently without prompting

* The descriptors shown are the “anchor” descriptors for scores 1, 3, and 5.
1.4.4 Procedure-Specific Assessment Scales

Procedure-specific assessment scales form the middle ground between checklists and global rating scales (Palter & Grantcharov, 2012b). They are specific enough to provide useful formative feedback to the individual that is being assessed; yet they are not as prescriptive as checklists. These scales are created in a similar manner to checklists, often using consensus of a group of experts to generate the items on the scale. Once a list of items is generated, each item is assigned subjective descriptive anchors of performance, often on a Likert-type scale. Some authors prefer to use the term “procedure-specific global rating scales” to describe procedure-specific assessment scales (R. Aggarwal, Grantcharov, Moorthy, Milland, & Darzi, 2008). Procedure-specific assessment scales have been developed for several procedures including laparoscopic colorectal surgery (Palter & Grantcharov, 2012b), laparoscopic right and sigmoid colectomy (Figure 8) (Sarker, Kumar, & Delaney, 2010), and endoscopy with and without polypectomy (Sarker, Albrani, Zaman, & Patel, 2008).

<table>
<thead>
<tr>
<th>Specific technical skills</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access &amp; port insertion</td>
<td>Created clumsily &amp; with difficulty</td>
<td>Created adequately</td>
<td>Created quickly &amp; skillfully</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retraction, dissection, exposure, &amp; mobilisation at ileocolic pedicle</td>
<td>Poor retraction, dissection, exposure, &amp; mobilisation at ileocolic pedicle, large amount of bleeding</td>
<td>Satisfactory retraction, dissection, exposure &amp; mobilisation at ileocolic pedicle, small amount of bleeding</td>
<td>Expert retraction, dissection, exposure &amp; mobilisation at ileocolic pedicle, minimal amount of bleeding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clipping &amp; transaction of ileocolic, right colic, &amp; middle colic arteries</td>
<td>Clips not placed accurately, large amount of bleeding</td>
<td>Clips placed proximally and distally adequately, minimal bleeding</td>
<td>Clips placed expertly, proximally and distally, with adequate with length, no bleeding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobilise colon along Toldt’s fascia on right side of colon</td>
<td>Poor mobilisation</td>
<td>Satisfactory mobilisation</td>
<td>Expert mobilisation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transection of colon</td>
<td>Clumsily done</td>
<td>Adequately done</td>
<td>Expertly &amp; smoothly done</td>
<td>Smoothly done with minimal trauma</td>
<td></td>
</tr>
<tr>
<td>Extraction of colonic specimen</td>
<td>Clumsily done</td>
<td>Adequately done with little trauma</td>
<td>Expertly &amp; smoothly done, good blood supply</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colonic anastomosis</td>
<td>Clumsily done, discrepancy in blood supply</td>
<td>Adequately done with little trauma, good blood supply</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Procedure-specific assessment scales have the advantages of both checklists and global rating scales. They provide formative feedback via identification of specific areas of weakness, while offering some flexibility in regards to the completion of specific operative steps.
1.4.5 Motion Analysis

Motion analysis offers another method for objective assessment of technical skills. It relies on motion tracking hardware and software to detect the motion of the hands or laparoscopic instruments. Software records several objective metrics, including distance travelled, speed and time to complete the task. Examples of motion tracking devices include the Imperial College Surgical Assessment Device (ICSAD; Imperial College, London, UK), the Advanced Dundee Psychomotor Tester (ADEPT; University of Dundee, Dundee, UK), the Hiroshima University Endoscopic Surgical Assessment Device (HUESAD; Hiroshima University, Hiroshima, Japan) and the TrEndo Tracking System (Delft University of Technology, Delft, The Netherlands) (van Hove et al., 2010).

There are several advantages of motion analysis systems. Performance metrics are obtained by the device and are completely objective. Presence of an experienced and trained preceptor is not required as the device does assessments automatically. These devices can be used in the laboratory and in the operating room. The disadvantages include the requirement for hardware and software resources, difficulties with using the metrics of distance, speed and time for formative assessment, and limitations of technology. For example, some devices are reliant on infrared tracking technology to detect the movement of the surgeon’s hands, which is then translated into the metrics of distance, speed and time by the machine. In the operating room, a member of a surgical team or a surgical instrument may block the infrared beam, which will result in the loss of data. Moreover, some surgeons may find it uncomfortable to operate with motion tracking sensors on their hands, which limits the feasibility of this type of assessment in the operating room.

1.4.6 Virtual Reality Simulators

Virtual reality (VR) simulators have been utilized to teach basic and intermediate laparoscopic skills for general surgery, gynecology, urology and other surgical specialties (Palter & Grantcharov, 2010). A number of different simulators are currently available on the market including the Minimally Invasive Surgical Trainer Virtual Reality (MIST-VR; Mentice, Gothenburg, Sweden), LapSim (Surgical Science, Gothenburg, Sweden), LAP Mentor (Simbionix Corporation, Cleveland, Ohio, USA) and others. All modern VR simulators have motion tracking devices imbedded into their hardware. As a result, these simulators form a very useful platform for objective assessment of technical skills. Most simulators will record simple
performance metrics of path length, economy of motion and time to task completion, which can in turn be used for objective assessment of technical skill.

The advantages of VR simulators are similar to motion tracking devices. There is no subjectivity in the assessment and there is no requirement to involve an experienced surgeon to assess the skills of a trainee. All assessments are done in a standardized and safe environment of a simulator, without confounding patient factors. The disadvantages include the inability to use this equipment in the operating room, the difficulty with using performance metrics for formative feedback and the expense associated with the initial purchase of a VR simulator.
1.5 Psychometric Properties of Assessment Scales

1.5.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


An ideal scale for assessment of technical and/or non-technical skills in surgery must be feasible to use, reliable, valid and generalizable.

1.5.2 Reliability

In psychometrics, reliability is defined as the overall consistency of a measure or an extent to which an experiment, test, or measuring procedure yields the same results on repeated trials (http://www.merriam-webster.com). Reliability can be sub-divided into internal consistency, inter-rater reliability and test-retest reliability. Internal consistency reflects the correlation between different items on an assessment scale and their contribution to the overall score on a scale (van Hove et al., 2010). Inter-rater reliability refers to the agreement between the scores of two or more raters evaluating the performance of the same individual (van Hove et al., 2010). Test-retest reliability refers to the agreement between the scores on an assessment scale for a task that was performed twice by the same individual. In the domain of surgery, performance of a surgeon on two consecutive operative cases with the same level of difficulty is used to assess the test-retest reliability.

1.5.3 Validity

Validity is defined as an ability of the assessment tool to measure a characteristic or a trait that it was designed to measure. Validity can be sub-divided into face, content, construct, concurrent
and predictive validity (van Hove et al., 2010). Face validity deals with the realism and functionality of the assessment tool. This concept is easier to grasp with an example. A model of bowel anastomosis with cadaveric porcine bowel would have high face validity, whereas a model with foam tubing would have low face validity. Content validity refers to the content of the items in an assessment scale and if those items are suited to measure the characteristic of interest. For example, an assessment scale that was designed to measure technical ability in suturing should contain items that address the task of suturing, such as loading the needle, pronation and supination of the hand. Construct validity is a measure of how well an assessment scale is measuring the construct of interest – an intangible collection of abstract concepts and principles that are inferred from behavior and explained by educational or psychological theory (Downing, 2003). For example, an assessment scale that purports to measure the construct of technical skill would have construct validity if it was able to discriminate between individuals with poor, intermediate and excellent technical skills. Concurrent validity refers to the agreement of the scores on the assessment scale of interest and a current “gold standard” for the assessment of that construct. For example, concurrent validity of a new assessment scale for technical skill could be demonstrated by correlating the scores on the new scale to the scores on the OSATS GRS, which is felt to be the current “gold standard” in the assessment of technical skill (van Hove et al., 2010). Lastly, predictive validity refers to the extent that a score on an assessment scale today would be predictive of performance in the future.

The concepts of face, content, construct, concurrent and predictive validity continue to dominate the surgical education literature; however, Cook and Beckman (Cook & Beckman, 2006) and Downing (Downing, 2003) have been advocating for a modification in the terminology and sources of validity in medical education research. Herein, I will review their proposed terminology; albeit, I will continue to use the concepts of face, content, construct, concurrent and predictive validity for the remainder of this thesis.

Downing, Cook and Beckman suggest that all validity should be looked at as construct validity, and that assessments should not be looked at as valid or invalid, rather as assessment scores having more (or less) validity evidence to support their proposed interpretations. They also suggest that construct validity requires multiple sources of evidence to support or refute a meaningful interpretation of an assessment score. These sources of evidence include: content, response process, internal structure, relationship to other variables and consequences. Downing provides a good example of content evidence for construct validity of a multiple-choice test
(Downing, 2003). Presence of a test blueprint, which shows a direct linkage of the questions on the test to the instructional objectives, is an example of content evidence. Similarly, other content evidence for construct validity includes evidence that item-writers for the multiple-choice test were qualified as content experts in that discipline, that there are sufficient numbers of questions to adequately sample the large content domain, and that test questions been edited for clarity.

Response process is defined as “evidence of data integrity such that all sources of error associated with the test administration are controlled or eliminated to the maximum extent possible” (Downing, 2003). An example of response process evidence for a multiple-choice examination would be documentation of the accuracy of the scoring key and documentation that poorly performing test items were removed from the final scoring. Internal structure relates to “the statistical or psychometric characteristics of the examination questions or performance prompts, the scale properties – such as reproducibility and generalizability, and the psychometric model used to score and scale the assessment” (Downing, 2003).

Evidence of relationship to other variables relates to the relationship of assessment scores to scores on some preexisting measure of the same construct. An example of evidence of relationship to other variables for a multiple-choice examination in medial school would be high correlation between scores on the multiple-choice examination and scores on an Objective Structured Clinical Encounter (OSCE) examination. Consequences refer to the impact of assessments on individuals and society at large.

1.5.4 Feasibility of Use and Generalizability

An idea assessment scale must be generalizable and feasible to use. A generalizable assessment scale can be used in multiple contexts. OSATS GRS, for example, has been used in a simulation laboratory and in the operating room to measure technical skills in general surgery, obstetrics and gynecology, urology and other procedure (van Hove et al., 2010).

Feasibility of an assessment scale refers to the time and resource requirements to use the scale. Dath and colleagues examined the feasibility of using an assessment scale to evaluate surgical performance for a low anterior resection and a Nissen fundoplication using video recordings (Dath et al., 2004). Ten blinded laparoscopic surgeons evaluated videos of 21 senior surgery residents performing the abovementioned operations in a pig. Investigators tried to minimize the rating time by allowing the raters to fast-forward through the video at their discretion. By being
able to fast-forwarding the video, the assessment time was shortened by 80% without a significant loss in inter-rater reliability. As assessment scale that is not feasible to use will have minimal uptake in practice.

1.5.5 Psychometric Properties of Available Assessment Scales

The reliability and validity of currently available scales for assessment of technical skills has been systematically reviewed Van Hove and colleagues (van Hove et al., 2010) and investigated by Aggarwal et al (R. Aggarwal et al., 2008) and Regehr et al (Regehr et al., 1998). Van Hove systematically reviewed 104 studies addressing the validity and reliability of scales for objective assessment of technical skills in surgery and gynecology. The level of evidence was assessed according to the Oxford Centre for Evidence-based Medicine levels of evidence. Twenty (19.2%) of studies had a level of evidence 1b or 2b. Twenty-eight studies (26.9%) reported on using an assessment scale in the operating room. Assessments were grouped into procedure-specific checklists, global rating scales, Objective Structured Assessment of Technical Skills, motion analysis, Virtual Reality simulators, video assessment and a miscellaneous category. The authors of the systematic review concluded the following:

(1) OSATS was most accepted as the ‘gold standard’ for objective skills assessment, albeit evidence for its use in the operating room was of a lesser grade and less abundant. OSATS was said to be well suited for use in formative assessment (feedback for improvement); however, it was not suitable for use in summative assessment (high-stakes “pass”/”fail” decisions).

(2) Evidence for reliability and validity of checklists and global rating scales other than OSATS GRS was limited. As a result, these methods could be used in formative assessment; however, use in summative assessment was not recommended.

(3) Motion analysis devices were suitable for use in formative but not summative assessment.

(4) Virtual reality simulators had acceptable reliability and construct validity. Cut-off values have been defined, thus, VR simulators could be used for summative assessment.
Videos can be used for formative assessment; however, there was insufficient evidence to suggest use of videos for summative assessment. Moreover, the authors suggested that editing, especially shortening videos, before assessment has a clear adverse impact on the outcome of assessment.

Assessment using the Fundamentals of Laparoscopic Surgery was reported to be well suited for formative assessment and, because cut-off values have also been defined, it could be used for summative assessment.

Aggarwal and colleagues examined feasibility, validity, inter-rater and test-retest reliability of the generic OSATS GRS, modified OSATS GRS, procedure-specific global rating scale and procedure-specific checklist for cholecystectomy (R. Aggarwal et al., 2008). Two experienced observers were recruited to review 47 operative cases of cholecystectomy performed by 6 novice and 13 experienced surgeons. Generic OSATS GRS and modified OSATS GRS had evidence of construct validity and a high degree of inter-rater and test-retest reliability. The procedure-specific global rating scale and the procedure-specific checklist for laparoscopic cholecystectomy did not show evidence of construct validity and had low to intermediate inter-rater and test-retest reliability.

Regehr and colleagues compared the psychometric properties of a global rating scale, a global rating scale proceeded by a checklist, and a checklist alone for surgery resident performance in an examination of technical skill using a type of Objective Structured Clinical Examination (Regehr et al., 1998). Fifty-three general surgery residents were recruited to take part in eight, 15-minute bench-top stations designed to test different surgical skills. Two non-blinded raters marked each station. One rater used the global rating scale and the other used a checklist followed by a global rating scale. Global rating scales scored by experts were shown to have greater inter-station reliability, better construct validity, and better concurrent validity than checklists. The presence of the checklists did not improve the reliability or validity of the global rating scale over that of the global rating scale alone. The authors concluded, “Global rating scales administered by experts are a more appropriate summative measure when assessing candidates on performance-based examinations”.

1.6 Feedback

1.6.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


Feedback is an important component in any educational intervention (Issenberg, McGaghie, Petrusa, Lee Gordon, & Scalese, 2005). Feedback is thought to produce motivation, supply reinforcement for correct actions, dissuade incorrect actions, and provide information about errors as a basis for corrections (Xeroulis et al., 2007). There are two major types of feedback: intrinsic and extrinsic (Magill, 2004). Intrinsic feedback relies on performance-related information that is available directly to the sensory system of the performer, such as visual or auditory perceptions during the performance of a task. Extrinsic feedback relies on performance-related information that is provided by an external source. The aim of the extrinsic feedback is to augment the intrinsic feedback (Magill, 2004). Extrinsic feedback has two sub-types: concurrent and summary (terminal). Concurrent feedback is administered during an intervention, whereas summary feedback is administered after an intervention.

The effects of concurrent and summary feedback on technical skill acquisition were compared in a prospective, single-blinded, randomized controlled trial that examined the effects of self-study with computer-based video instruction, expert feedback during practice trials (concurrent), and expert feedback after practice trials (summary) on acquisition and retention (1 month after completion of training) of basic surgical skills (Xeroulis et al., 2007). Sixty first-year medical students were randomized into 4 groups (no feedback, computer-based video instruction feedback, concurrent feedback and summary feedback) and were taught simple interrupted suturing in a simulation laboratory. Performance was videotaped and assessed by 2-blinded reviewers using a global rating scale and motion analysis. At the initial post-test, there were no
differences in performance between computer-based video instruction group, concurrent feedback group, and summary feedback group. All three feedback groups performed better than the no feedback group. At a 1-month retention test, computer-based video instruction group and summary feedback group outperformed the concurrent feedback and no feedback groups. The authors concluded, “summary feedback is more effective that concurrent feedback in teaching basic technical skills to medical students”.

The aforementioned findings for a suturing task were replicated for an endoscopy task (Walsh, Ling, Wang, & Carnahan, 2009). Thirty novice endoscopists were randomized to either concurrent or summary feedback and were asked to complete 12 colonoscopies in a simulated environment. Performance was assessed on a pre-test, a post-test, a retention test (1-week) and a transfer test (a different colonoscopy task in a simulated environment). Two independent raters assessed participant’s performance using a checklist, a global rating scale and the metric of time. Both groups demonstrated equivalent performance on a pre-test, a post-test and a retention test. The summary feedback group outperformed the concurrent feedback group on the transfer test. The authors of the study suggested that terminal feedback was more effective in teaching basic endoscopic procedures than concurrent feedback.

The impact of intense versus limited (less than 10 minutes) instructor-led summary feedback on acquisition of laparoscopic suturing skills in novice medical students was investigated in a study that compared 3 groups of trainees: Group I (one video-based tutorial with intense instructor feedback), Group II (one video-based tutorial with limited instructor feedback) and Group III (multiple video-based tutorials with limited instructor feedback) (Stefanidis, Korndorffer, Heniford, & Scott, 2007). Group II required less time and fewer repetitions on a simulator to achieve proficiency in comparison to Group I. The cost of training in Group II was also less than in Group I. As a result, limited summary feedback appeared to be superior to intense summary feedback. Furthermore, intense feedback during early stages of skill acquisition was hypothesized to hinder learning.

The beneficial effects of summary feedback can be explained using the Guidance Hypothesis proposed by Winstein and Schmidt (Winstein & Schmidt, 1990). In this hypothesis, feedback is expected to guide the learner toward the intended goal by providing information on error correction and by keeping the learner motivated and interested. If feedback is too frequent as is the case with concurrent feedback, the learner may develop a dependency on this type of
feedback, and its presence may become essential for performance. When concurrent feedback is removed as in the case with retention testing, the learner’s performance suffers. The guidance hypothesis suggests that concurrent feedback may direct a learner’s attention away from the critical intrinsic feedback, whereas summary feedback may allow a trainee to process and consolidate intrinsic feedback during the task and receive extrinsic feedback at the end of the task.
1.7 Surgical Simulation and Non-Technical Skills Training

1.7.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language

Exclusions: Publications within the disciplines of anesthesia and emergency medicine


The operating room is a unique environment where professionals from multiple specialties are required to work together in a closely coordinated fashion. Non-technical skills, such as communication, teamwork, situation awareness and leadership, can have a profound impact on that unique working environment. Failures in non-technical skills may lead to an increased number of errors in surgery (Shouhed, Gewertz, Wiegmann, & Catchpole, 2012). Incomplete or erroneous communication was reported as the causal factor in 43% of errors made during an operation (Gawande, Zinner, Studdert, & Brennan, 2003). Thirty-six percent of communication errors in the operating room were shown to result in inefficiency, team tension, waste of resources, patient inconvenience and procedural error (Lingard et al., 2008). Participation in team-based training courses has been associated with improved team skills, better satisfaction with care, improved compliance with pre-operative briefings and decreased error rate in the operating room (Shouhed et al., 2012).

Poor non-technical skills have been shown to impact technical skills in the operating room (Hull et al., 2012). Increased situation awareness has been linked to a decrease in errors in the operating room, whereas failures in teamwork have been associated with an increase in errors.
A recent systematic review of the impact of non-technical skills on technical performance in surgery concluded that failures in non-technical skills were associated with a greater rate of technical errors in the operating room (Hull et al., 2012). The importance of non-technical skills in surgery has been highlighted in the recent recommendations of the Accreditation Council for Graduate Medical Education in the United States for surgery trainees to demonstrate mastery of teamwork-related competencies (ACGME, 2012; Sanfey, McDowell, Meier, & Dunnington, 2011). To address the development of non-technical skills in surgery, the American College of Surgeons (ACS) and the Association for Program Directors in Surgery (APDS) have developed Phase 3 of the Surgical Skills Curriculum for Residents; and the United States Department of Defense Patient Safety Program in collaboration with the Agency for Healthcare Research and Quality developed the Team Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS) curriculum (http://teamstepps.ahrq.gov).

Simulation centers offer a great environment to teach and evaluate non-technical skills (Arora, Miskovic, et al., 2011; Black, Nestel, Kneebone, & Wolfe, 2010). High fidelity, simulated operating rooms have already been used for collaborative training of surgical teams (Paige et al., 2009), whereas, hybrid simulations have been used to combine the training of technical and non-technical skills (Gettman et al., 2009). A recent systematic review of the current approaches to and outcomes of non-technical skills training provided an excellent overview of current approaches and interventions aimed at improving patient-centered communication, teamwork, surgical decision making, ability to cope with stress, patient safety and error reduction in surgery trainees (Dedy, Bonrath, Zevin, & Grantcharov, 2013). The following conclusions were drawn from the available evidence:

1. Basic communication skills can be effectively taught to residents by means of simulated patient encounters in conjunction with structured, formative feedback. Combining patient communication scenarios with basic procedural tasks in standardized modules may allow for time-efficient training and assessment of technical and non-technical skills within surgical curricula.

2. High-fidelity crisis simulations followed by debriefing or feedback sessions were the most commonly used instructional strategies for team training. A simulation-based approach to team training was shown to be superior to a purely didactic approach. Performance assessments were shown to be important for provision of structured
feedback in debriefing sessions, thereby allowing trainees to reflect on their behavior and remediate their mistakes.

3. Designated non-technical skills training comprised of video examples, group discussions, and hands-on practice was postulated to result in improved decision-making during surgical procedures.

4. Patient safety and error management strategies including pre-operative briefings and checklists have been implemented in multiple healthcare facilities and have shown medium-term decreases in perioperative morbidity and mortality. Instruction on the correct use of these strategies should be integrated into surgical curricula via simulation-based or role-play briefing exercises, followed a feedback session with discussions about the relevance of safety checklists and briefing.

There are a number of different objective scales for assessment of non-technical skills in surgery. Examples of such scales include the Non-technical Skills for Surgeons (NOTSS) (Yule et al., 2006), the Oxford Non-technical Skills (NOTECHS) presented in Figure 9 (Mishra, Catchpole, & McCulloch, 2009) and the Observational Teamwork Assessment for Surgery (OTAS) (Sevdalis et al., 2009). The uses, advantages, disadvantages and psychometric properties of each scale have been well summarized by Sharma and colleagues (Sharma, Mishra, Aggarwal, & Grantcharov, 2011).
1.8 Theoretical Frameworks for Acquisition of Technical Skills

1.8.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


Two theories in medical education are useful for theoretical understanding of how technical and non-technical skills are acquired in surgery: the motor learning theory and the cognitive load theory.

1.8.2 Motor Learning Theory

The motor learning theory was proposed by Fitts and Posner in 1967 and has been accepted in both the psychological literature as well as the medical education literature (Fitts & Posner, 1967). Reznick and MacRae have summarized the application of this theory to the acquisition of technical skills in surgery (Reznick & MacRae, 2006). The cognitive stage is the first stage in the motor learning theory. During this stage the learner begins to intellectualize the technical procedure at hand. His or her performance is erratic, and the procedure is carried out in sequential, distinct steps. With continued practice and appropriate feedback, the learner will move from the cognitive to the integrative stage. In the integrative stage, the learner begins to translate the knowledge acquired in the cognitive stage into appropriate motor behaviors. Tasks are executed with more fluidity and with fewer interruptions. With ongoing practice, the learner transitions from the integrative to the autonomous stage, where motor performance becomes more smooth and automated. At this stage the learner no longer needs to think about how to execute a particular procedure and can instead concentrate on other aspects of patient care.
I will use the task of laparoscopic suturing to illustrate the progression of a novice learner from the cognitive to the autonomous stage of the motor learning theory. In the cognitive stage, the learner should spend time learning about laparoscopic instruments, depth perception in minimally invasive surgery, 2D to 3D conversion, different types of sutures and needles, the importance of pronation and supination of the wrist, and the appropriate position of the needle in the needle driver. Once this knowledge is acquired in the cognitive stage, the learner begins to practice laparoscopic suturing and progresses into the integrative stage. Initially the learner has to learn how to load the needle correctly, how to pronate the wrist and how to enter the tissue at 90 degrees. With appropriate feedback from an instructor, the learner begins to move through the motions with greater fluidly and with fewer pauses. With ongoing practice and appropriate feedback, movements become more efficient, faster and safer. The learner progresses to the autonomous stage of skill acquisition. He or she no longer has to think about how to load the needle or to remember to pronate the hand. Task execution becomes automatic.

1.8.3 Cognitive Load Theory

The cognitive load theory was proposed by Sweller in 1980’s (Sweller, 1988), and its role in health professional education was recently discussed by van Merrienboer and Sweller (van Merrienboer & Sweller, 2010). The main assumption of the cognitive load theory is that the human cognitive system has a limited working memory that can hold no more than five to nine information elements and can process no more than two to four elements at a time. Our working memory is able to deal with these information elements for only a couple of seconds, and most of the elements are lost after approximately 20 seconds unless refreshed by rehearsal. It is important to note that these limitations of the working memory apply only to novel information obtained through sensory memory and not to the information retrieved from long-term memory. Long-term memory is hypothesized to hold cognitive schemas of various degrees of complexity and automation; organization of knowledge into these schemas contributes to human expertise (van Merrienboer & Sweller, 2010). With extensive practice, learners combine multiple simple ideas into complex schemas, which are stored in long-term memory.

The cognitive load theory breaks down cognitive load into intrinsic load, extraneous load and germane load. Intrinsic load is generated from the intrinsic nature of learning the task, and thus, cannot be altered by instructional interventions without altering the task to be learned (simplification of the task) or by the act of learning itself. Extraneous load is generated by the instructional procedures. Poor guidance during task completion results in increased extraneous
load. Germane load refers to the working memory resources used to deal with the intrinsic cognitive load that lead to learning. The cognitive load theory assumes that intrinsic and extraneous cognitive load are additive. As a result, the sum of intrinsic and extraneous loads for basic surgical tasks is unlikely to result in cognitive overload; however, the sum of intrinsic and extraneous loads for complex tasks such as laparoscopic surgical procedures may easily surpass the working memory capacity and yield overload (Figure 10). Strategies that target the extraneous and intrinsic load are expected to prevent cognitive overload. Extraneous load can be managed by training in a simulated environment, where multiple sources of information are replaced by a single source – a surgical simulator. Intrinsic load can be decreased by separating complex tasks into a series of sub-tasks, which can be performed in a low-fidelity environment. With increased practice in a low-fidelity simulated environment, cognitive schemas will be developed in long-term memory, and a trainee will be able to progress to increasingly higher-fidelity environments.

Figure 10. The additive effects of intrinsic and extraneous cognitive load: (a) overload; (b) preventing overload by decreasing extraneous load (adopted from van Merrienboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ. Jan 2010;44(1):85-93).
1.9 Transfer of Skill from the Laboratory to the Operating Room

1.9.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


The ultimate confirmation of the effectiveness of any simulation intervention is the demonstration of skill transfer from a simulated environment into a clinical setting in the real world. A substantial body of evidence supports the concept of transfer of technical and non-technical skills from a simulated environment to the operating room (R. Aggarwal, Ward, et al., 2007; Ahlberg et al., 2007a; Grantcharov et al., 2004a; Korndorffer et al., 2005; Sturm et al., 2008; Zendejas et al., 2013). Herein, I will highlight some of this evidence.

Grantcharov and colleagues carried out a randomized controlled trial to compare the effect of proficiency-based training on a virtual reality simulator on technical performance in the operating room (Grantcharov et al., 2004a). Twenty surgery trainees with limited laparoscopic experience were randomized to either the VR training group or the control group (no training). All participants performed one laparoscopic cholecystectomy for baseline assessment of their laparoscopic skills. Following baseline assessment, the VR group trained to proficiency on a VR simulator, while the control group received no training. Fourteen days after enrolment in the study, all participants performed one laparoscopic cholecystectomy in the operating room. Participants in the VR group were significantly faster, committed fewer errors, and demonstrated greater economy of movement in the operating room compared to participants in the control group.
Aggarwal and colleagues built on the work of Grantcharov et al. and examined the effects of simulation-based training on the learning curve for a basic laparoscopic operation (R. Aggarwal, Ward, et al., 2007). In their study, 20 novice surgeons were randomly allocated to a VR-trained group or a conventionally trained group. The VR-trained group completed a proficiency-based curriculum on a VR simulator followed by three laparoscopic cholecystectomies using cadaveric porcine liver. The conventionally trained group performed five laparoscopic cholecystectomies on the cadaveric porcine liver. Technical performance of novice surgeons after training was compared to the performance of experienced surgeons. Technical skill was assessed using motion analysis and a video-based global rating scale. The VR-trained group was significantly faster, performed significantly fewer movements and achieved significantly greater global rating scale scores on their first laparoscopic cholecystectomy as compared to the conventionally trained group. Furthermore, statistical equivalence of performance between the VR-trained group and the conventionally trained group was achieved on the 3<sup>rd</sup> cholecystectomy for the VR group and 5<sup>th</sup> cholecystectomy for the conventionally trained group. These results suggested that simulation-based training lessened the learning curve in the operating room.

Systematic reviews by Strum et al (Sturm et al., 2008) and Zendejas et al (Zendejas et al., 2013) reported similar results for the transfer of technical skills. Strum’s review included 10 randomized controlled trials and one nonrandomized comparative trial. Outcome measures were overall performance, performance time, ability to complete procedure, senior supervising surgeon takeover, performance errors, flow of procedure, time and motion, staff productivity, patient discomfort and confidence after simulation-based training. Based on the available evidence, simulation-based training appeared to result in transfer of skills to the operating room setting. However, the authors were careful to point out that the evidence base was relatively weak with more than half of the randomized controlled trials not reporting methods of randomization, allocation concealment, intention to treat analysis, losses to assessment, study period and exclusion criteria. Sample sizes were also generally small. The authors concluded, “More studies are required to strengthen the evidence base and to provide the evidence needed to determine the extent to which simulation should become a part of surgical training programs.”

Zendejas and colleagues built on Strum’s work by systematically reviewing the literature through May 2011 and conducting a meta-analysis of 219 studies (91 randomized controlled trials) that compared simulation-based training for laparoscopic surgery in comparison to no intervention or an alternate training activity (Zendejas et al., 2013). Outcomes included satisfaction, skills (in a
test setting) of time (to perform the task), process (eg. performance rating), product (eg. knot strength), behaviors when caring for patients and patient effects. In comparison with no intervention, pooled effect sizes (ES) favored simulation for outcomes of knowledge (ES=1.18), skill time (ES=1.13), skill process (ES=1.23), skill product (ES=1.09), behavior time (ES=1.15), behavior process (ES=1.22) and patient effects (ES=1.28). When compared with non‐simulation-based instruction, results favored simulation-based training for outcomes of skill time (ES=0.75) and skill process (ES=0.54). The authors concluded, “Simulation-based laparoscopic surgery training of health professionals has large benefits when compared with no intervention and is moderately more effective than non-simulation instruction”.

1.10 Simulation-Enhanced Training and Patient Outcomes

1.10.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


Despite the convincing evidence that simulation-enhanced training is associated with large effects for the outcomes of knowledge, skills and behaviors (Cook et al., 2011), there is still a paucity of studies in surgery linking simulation-enhanced training to improved patient outcomes. In a recent systematic review and meta-analysis of technology-enhanced simulation in the education of health professions, Cook and colleagues identified 32 studies with 1648 participants that reported on the effects of simulation-enhanced training on patient outcomes (Cook et al., 2011). For those studies that compared simulation-enhanced training with no intervention, simulation-enhanced training was associated with a moderate pooled effect size of 0.50 (95% CI, 0.34-0.66; $P=0.001$) with high inconsistency and range of effect sizes from -0.28 to 1.68. For the subgroup of 14 randomized controlled trials, the effect size was slightly less at 0.37. Interestingly, there was only one study that examined the effects of simulation-enhanced training on patient outcome in surgery (Zendejas et al., 2011). In that study, 50 general surgery residents were randomly assigned to either the mastery-learning curriculum (cognitive training using web-based modules and proficiency-based training on a total extraperitoneal inguinal hernia simulator) or the standard general surgery practice. Participants in the standard practice arm were asked to perform at least one laparoscopic inguinal hernia repair under supervision in the operating room. Participants in the mastery-learning curriculum were required to pass the curriculum prior to competing at least one laparoscopic inguinal hernia repair in the operating room. Each participant performed at least one laparoscopic inguinal hernia repair in the operating
room. The outcomes of interest were operative time, operative performance, proportion of procedure performed by the trainee, intraoperative complications (vessel, bladder or bowel injury, peritoneal tear, etc.), postoperative complications (hematoma, seroma, skin infection, urinary retention), need for an overnight stay, recurrence of the inguinal hernia and chronic groin pain. The mastery-learning group had significantly fewer intraoperative complications, postoperative complications and overnight stays compared to the standard practice group. With a median follow-up of 5.2 months (range 0–12), the number of patients who experienced a hernia recurrence or were evaluated for groin pain at least 3 months post repair were similar between the groups.
1.11 Cost-Effectiveness of Simulation-Enhanced Training

1.11.1 Search Strategy

Databases: MEDLINE (until April 2013), EMBASE (until April 2013), Scopus (until April 2013), ERIC (until April 2013)

Limits: English language


Despite strong evidence of knowledge and skill transfer from the simulated to the clinical environment, there is still a lack of wide implementation of simulation-enhanced training across surgery training programs. The paucity of cost-effectiveness analyses of simulation-enhanced training is one factor that is contributing to this lack of implementation. Simulation-enhanced training is expensive, but so is conventional training. Bridges and Diamond reported $47,970 USD as the cost of training one surgery resident for 4 years in the operating room (Bridges & Diamond, 1999). Harrington and colleagues reported a similar figure of $45,061 USD per year as the cost of training 15 senior residents to perform 2 laparoscopic entero-enterostomies (Harrington et al., 2007). In a recent systematic review of 967 comparative studies on simulation-enhanced training, only 59 studies (6.1%) reported some cost elements and only 15 (1.6%) provided information on costs compared with another instructional approach (Zendejas, Wang, Brydges, Hamstra, & Cook, 2012). Of these 15 studies, only 4 compared simulated-enhanced training with another instructional modality, and none of these studies were conducted in a surgery setting. Out of these 4 studies, two showed that simulation costs more and is educationally more effective (de Giovanni, Roberts, & Norman, 2009; Petscavage et al., 2011), one study reported that simulation was more costly and similarly as effective (Delasobera et al., 2010), and one study showed simulation to be less costly and more effective (Limpaphayom, Ajello, Reinprayoon, Lumbiganon, & Gaffikin, 1997).

Orzech and colleagues conducted a prospective, single-blinded, randomized trial, which allocated 24 surgery residents to proficiency-based training on a VR simulator or a box trainer
(Orzech et al., 2012). Outcome measures of interest were the technical skill scores for placement of intracorporeal laparoscopic stitches during a Nissen fundoplication in the OR and the cost of training. Analysis of cost for training using a box trainer, a VR simulator and a conventional training program revealed that the annual cost of training 5 residents on a box trainer was $11,975.00 CAD per year; on the VR simulator was $77,500.00 CAD per year; and in conventional residency was $17,380.00 CAD per year. Over the span of 5 years, box training was reported to be the most cost-effective option. The main limitation of this study was the failure to account for such ancillary costs as rental of simulation space, presence of a simulation technician, maintenance costs, etc. This study utilized technical skill scores to compare performance in the operating room, while patient outcomes were not recorded. Therefore, the true cost-effectiveness of simulation-enhanced training from a patient care perspective remains unknown. Additional studies are required to answer the question of whether simulation-enhanced training is more cost-effective than conventional training in basic, intermediate and advanced laparoscopic operations.
1.12 Simulation-Based Training and Learning Curves in Laparoscopic Bariatric Surgery

This section is modified from the following:


1.12.1 Abstract

**Background:** Ex-vivo simulation-based technical skills training has been shown to improve operating room performance and shorten learning curves for basic laparoscopic procedures. The application of such training for laparoscopic Roux-en-Y gastric bypass (LRYGB) has not been reviewed.

**Methods:** Relevant studies were identified by one author from a search of MEDLINE and EMBASE databases from January 1, 1994 to November 30, 2010. Studies examining the learning curves and ex-vivo training methods for LRYGBP were included; all other types of bariatric operations were excluded. Manual search of the references was also performed to identify additional potentially relevant papers.

**Results:** Twelve studies (5 prospective and 7 retrospective case series) were selected for review. The learning curve for LRYGB was reported to be 50-100 cases. Bench-top laparoscopic jejunojejunostomy, anesthetized animals and Thiel human cadavers made up the bulk of the reported models for ex-vivo training. Most studies were of relatively poor quality. An evidence-based ex-vivo training curriculum for LRYGB is currently lacking.

**Conclusions:** Better quality studies are needed to define the learning curve for LRYGB. Future studies should focus on the design and validation of training models and a comprehensive curriculum for training and assessment of cognitive, technical and non-technical components of competency for laparoscopic bariatric surgery.
1.12.2 Introduction

The laparoscopic approach to bariatric surgery is currently considered the “gold standard” for surgical management of obesity (Nguyen et al., 2001; P. R. Schauer & Ikramuddin, 2001), and laparoscopic Roux-en-Y gastric bypass (LRYGB) is the most commonly performed bariatric operation (Livingston, 2010; Seki & Kasama, 2010). The minimally invasive approach to bariatric surgery is equivalent to laparotomy in rates of mortality, weight loss and reduction of co-morbidities (Nguyen et al., 2001; Puzziferri, Austrheim-Smith, Wolfe, Wilson, & Nguyen, 2006). LRYGB has the added benefit of decreased rates of surgical site infections, incisional hernias, intraoperative blood loss, length of stay, postoperative pain and respiratory complications, as well as better cosmesis (Nguyen et al., 2001; Puzziferri et al., 2006).

Training in laparoscopic bariatric surgery and LRYGB continues to be a challenge for surgery trainees and practicing surgeons alike. This challenge is a result of the long instruments which fulcrum at the abdominal wall, the counterintuitive visual fields, the decrease in tactile sensation and conduct of the operation in all quadrants of the abdomen (P. R. Schauer & Ikramuddin, 2001). In addition, the need for intracorporeal suturing and the possibility of morbidity and mortality results in a longer learning curve for the minimally invasive approach.

Simulation training shortens the learning curve for basic laparoscopic operations (R. Aggarwal, Ward, et al., 2007). A number of simulation-based models are available for teaching minimally invasive technical skills outside the operating room. These include video-box trainers (Vassiliou, Dunkin, Marks, & Fried, 2010), virtual reality simulators (Panait et al., 2011), synthetic models (S. M. B. I. Botden, L. Christie, R. Goossens, & J. J. Jakimowicz, 2010), live animal models, cadaveric animal organs and human cadavers (Levine et al., 2006). Training on some of these models has been shown to translate into improved technical performance in the operating room (Kanumuri et al., 2008; Levine et al., 2006; Van Sickle et al., 2008).

Unfortunately, very few ex-vivo training models for LRYGB have been developed (R. Aggarwal, Boza, et al., 2007) and no one has implemented these in clinical practice thus far.

This review aimed (1) to define the learning curve for LRYGB, (2) to examine current ex-vivo training strategies to shorten that learning curve, and (3) to propose a structure for a comprehensive simulation-based training curriculum for laparoscopic bariatric surgery.
1.12.3 Methods

1.12.3.1 Search strategy

One reviewer (BZ) searched MEDLINE and EMBASE databases from January 1, 1994 to November 30, 2010 using the following MeSH terms and keywords: “curriculum”, “general surgery”, “bariatric surgery”, “laparoscopy”, “internship and residency”, “models, educational”, “teaching”, “education”, “education, medical”, “surgical education” and “learning curve”. Relevant titles were identified, abstracts were read by one author (BZ) and eligible articles were retrieved. A manual cross-reference search of the bibliographies of retrieved articles was performed to identify additional studies.

1.12.3.2 Definitions

For the purpose of this review, the learning curve was defined as the number of cases a surgeon needs to perform to reach a plateau in operating times, conversion rates, complications and mortality.

1.12.3.3 Inclusion and exclusion criteria

Included studies had to report on learning curves and ex-vivo training models for LRYGB. Studies on other laparoscopic bariatric operations were excluded.

1.12.3.4 Data collection process

The following data were extracted by one author (BZ) from each study: study design; learning curve (number of cases); means and standard deviations of operating time (pre and post learning curve); percentage and types of complications (pre and post learning curve); and length of stay in the hospital (pre and post learning curve).

1.12.3.5 Data synthesis

The number of cases required to overcome the learning curve, as well as the mean and standard deviation (SD) of the operating time before and after the learning curve were abstracted from each study and imported into a Microsoft Excel worksheet (Microsoft Corporation, Redmond, WA, USA). The differences between means and their SD was then calculated. For studies that did not report a SD, it was not possible to calculate the SD of the difference. The overall mean (weighted mean) of the differences was then calculated using the formula
Overall\text{Mean} = \frac{\sum nD}{\sum n}

where (D) is the difference between the means for each study and (n) is the number of cases to overcome the learning curve for each study.
1.12.4 Results & Discussion

1.12.4.1 Defining the learning curve for laparoscopic Roux-en-Y gastric bypass

A total of 12 studies (C. G. Andrew, W. Hanna, D. Look, A. P. H. McLean, & N. V. Christou, 2006; Huang, Lee, Hung, Chen, & Tai, 2008; Kligman, Thomas, & Saxe, 2003; Lublin et al., 2005; Oliak et al., 2003; Pournaras et al., 2010; Schaeffer, Rusnak, & Amson, 2008; P. Schauer, Ikramuddin, Hamad, & Gourash, 2003; Shikora, Kim, Tarnoff, Raskin, & Shore, 2005; Shin, 2005; Sovik et al., 2009; Suter, Giusti, Heraief, Zysset, & Calmes, 2003) on the learning curve for the LRYGB were selected for review. Five studies (Lublin et al., 2005; Pournaras et al., 2010; Schaeffer et al., 2008; Shikora et al., 2005; Suter et al., 2003) were prospective case series and seven studies (C. G. Andrew, W. Hanna, D. Look, A. P. McLean, & N. V. Christou, 2006; Huang et al., 2008; Kligman et al., 2003; Oliak et al., 2003; P. Schauer et al., 2003; Shin, 2005; Sovik et al., 2009) were retrospective case series (Table 1 and 2).

The learning curve for the laparoscopic Roux-en-Y gastric bypass was reported to range from 50 to 100 cases. This finding is in agreement with the recommendations from the American Society for Metabolic and Bariatric Surgery (ASMBS) and the Fellowship Council, which mandate completion of a minimum of 100 laparoscopic bariatric cases during the fellowship with 51 completed as a primary surgeon. Greater numbers of cases were noted for surgeons with no prior laparoscopic experience and insufficient proctorship along the learning curve. This long learning curve is likely a result of the technical complexity of the laparoscopic approach, as stipulated earlier. Complications decreased from 9 – 42% early on in a surgeon’s learning curve to 0 – 22% after completion of the learning curve (Table 1 and 2). The duration of operation and rate of complications were the only two outcome variables that consistently decreased as the surgeon progressed along the learning curve. The estimated decrease in time to complete the LRYGB before and after the learning curve was 69 ± 39 min (Figure 10). A plateau in time to complete the operation was often seen only after 400 cases, which supports the concept of ongoing skill acquisition beyond the 100 case limit (Shin, 2005).
Table 1. Summary of retrospective case series defining the learning curve for LRYGB.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Learning curve (no. of procedures)</th>
<th>Duration of operation (min)*</th>
<th>Complications</th>
<th>Length of stay (days)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Before</td>
<td>After</td>
<td>Before (%)</td>
</tr>
<tr>
<td>Andrew et al.</td>
<td>2006</td>
<td>70-75</td>
<td>145(35)</td>
<td>119(23)</td>
</tr>
<tr>
<td>Huang et al.</td>
<td>2008</td>
<td>50</td>
<td>216(50-90)</td>
<td>105-2(57-8)</td>
</tr>
<tr>
<td>Klipman et al.</td>
<td>2003</td>
<td>NA</td>
<td>324(124)</td>
<td>168(40)</td>
</tr>
<tr>
<td>Ollak et al.</td>
<td>2003</td>
<td>75</td>
<td>189</td>
<td>125</td>
</tr>
<tr>
<td>Schauer et al.</td>
<td>2003</td>
<td>100</td>
<td>311</td>
<td>237</td>
</tr>
<tr>
<td>Shin</td>
<td>2005</td>
<td>50</td>
<td>113</td>
<td>73</td>
</tr>
<tr>
<td>Sevik et al.</td>
<td>2009</td>
<td>100</td>
<td>164(75)</td>
<td>68(21)</td>
</tr>
</tbody>
</table>

*Values are mean(s.d.) unless indicated otherwise; †values are median (range). Values are shown for before and after the learning curve. Overall (weighted) change in mean(s.d.) operating time following completion of predefined learning curve: 83(44) min. GJ, gastrojejunostomy; NA, data not available; GI, gastrointestinal; JJ, jejunojunostomy.

Table 2. Summary of prospective case series defining the learning curve for LRYGB.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Learning curve (no. of procedures)</th>
<th>Duration of operation (min)*</th>
<th>Complications</th>
<th>Length of stay (days)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Before</td>
<td>After</td>
<td>Before (%)</td>
</tr>
<tr>
<td>Lublin et al.</td>
<td>2005</td>
<td>80</td>
<td>249(70)</td>
<td>183(42)</td>
</tr>
<tr>
<td>Pournaras et al.</td>
<td>2010</td>
<td>100</td>
<td>162-8(53-0)</td>
<td>118-9(37-2)</td>
</tr>
<tr>
<td>Schaefer et al.</td>
<td>2008</td>
<td>&gt; 60</td>
<td>173(24)</td>
<td>145(22)</td>
</tr>
<tr>
<td>Shikora et al.</td>
<td>2005</td>
<td>100</td>
<td>212</td>
<td>163</td>
</tr>
<tr>
<td>Sultar et al.</td>
<td>2003</td>
<td>&gt; 100</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Values are mean(s.d.). Values are shown for before and after the learning curve. Overall (weighted) change in mean(s.d.) operating time following completion of predefined learning curve: 47(15) min. NA, data not available; RYGB, Roux-en-Y gastric bypass; GI, gastrointestinal; GJ, gastrojejunostomy; JJ, jejunojunostomy.
The major limitation of most studies in this review is their retrospective study design (Table 1) and the use of operative time as a surrogate for ascendance of the learning curve. A case series design divides patients into arbitrary groups based on case sequence (early vs. late) and then compares operative times and complication rates. Diminishing operative time may reflect the ascendance of the learning curve. However, the learning curve will also vary substantially depending on the assistant’s experience and the work environment including the experience of allied health personnel. Future prospective studies with independent external evaluation are recommended to better assess the learning curves for LRYGB.

Prior to the formal introduction of bariatric surgery fellowships, practicing surgeons often learned new techniques by visiting experienced colleagues for several days or by being proctored at their own institution for a number of cases. The finding of greater numbers of cases to overcome the learning curve for surgeons with no prior laparoscopic experience and insufficient
proctorship along the learning curve is not unexpected. Surgeons who are trying to learn laparoscopic bariatric surgery outside of a formal structured training program with appropriate formative feedback, opportunities for practice and mentorship, are expected to require greater number of cases to gain the same proficiency as those surgeons who get formal, structured training. The results of the study by Oliak and colleagues support the abovementioned statement (Oliak, Owens, & Schmidt, 2004). Morbidity, mortality and operative time for 75 consecutive LRYGB cases operated on by two laparoscopic surgeons, one with laparoscopic gastric bypass fellowship training (Group A) and one without laparoscopic bypass fellowship training (Group B) were compared. Operative time was significantly longer in Group B (189 min. vs 122 min., \(P<0.05\)). Major complications occurred more frequently in Group B (13% vs 8%, \(P=NS\)). In addition, the complications in Group B were more severe, resulting in 2 deaths. No deaths occurred in Group A.
1.12.4.2 Strategies for ex-vivo training in LRYGB

In the 21st century, simulation-based training has become much more ubiquitous in surgical education. In the era of outcome-driven health care and high costs of teaching in the operating room, ex-vivo training methods have been hypothesized to increase patients’ safety along a surgeon’s learning curve (Palter & Grantcharov, 2010). General surgery training programs in the United States now mandate completion of a basic laparoscopic training curriculum (Fundamentals of Laparoscopic Surgery) in a laboratory environment before Board of Surgery examination (Soper & Fried, 2008).

The advantages of simulation-based training include greater patient safety, decreased operating room costs (Harrington et al., 2007), improved operating room performance (Grantcharov et al., 2004b; Sroka et al., 2010) and superior cognitive learning (Palter, Grantcharov, et al., 2011). Unfortunately, objective curricula for simulation-based training exist only for basic laparoscopic skills (Derossis et al., 1998; Panait, Bell, Roberts, & Duffy, 2008) and one basic laparoscopic procedure (R. Aggarwal et al., 2009a). To date, there are no simulation-based curricula for advanced minimally invasive procedures such as the laparoscopic Roux-en-Y gastric bypass.

Currently, most of the training in laparoscopic bariatric surgery and laparoscopic Roux-en-Y gastric bypass occurs in the operating room. Adequate exposure to laparoscopic bariatric surgery is often achieved by completing a minimally invasive and/or bariatric surgery fellowship. The difficulty with training surgery residents in laparoscopic Roux-en-Y gastric bypass arises from the requirement for advanced laparoscopic skills (M. J. Martin, Eckert, Eggebroten, & Beekley, 2010). Naturally, proficiency in these skills is often only achieved by the trainee’s senior or chief year. These trainees can become primary assistants and surgeons in laparoscopic bariatric surgery if given an adequate caseload and by developing specific technical skills beforehand, for instance by a Forward Chaining Training Model (Peck & Detweiler, 2000; Rovito, Kreitz, Harrison, Miller, & Shimer, 2005). The Forward Chaining Training Model allows a trainee to learn a complex multistep task by performing each individual step in its natural sequence until the entire task is mastered (Peck & Detweiler, 2000). An application of this training model to the LRYGB was reported by Lublin and colleagues (Lublin et al., 2005).

If trainees can achieve proficiency in advanced laparoscopic skills at an earlier stage of their training, they will likely have more opportunities to participate in laparoscopic bariatric surgery. However, it is probably not possible for trainees to achieve proficiency during a structured 2-3
months rotation in bariatric surgery, because this is not enough to overcome the learning curve. Learning advanced laparoscopic techniques in a simulation laboratory prior to a rotation in bariatric surgery may provide the residents with skills to get them over the early part of the learning curve. This training results in a “pre-trained novice” - a trainee who has trained using simulation to a point where many of the psychomotor skills and spatial judgments are automated (Gallagher et al., 2005). Once in the operating room, such a trainee can focus on higher-level skills, such as learning the steps of the operation and management of complications. Therefore, junior surgery residents should practice in a simulated environment until basic psychomotor tasks are automated (Gallagher et al., 2005).

1.12.4.2.1 Ex-vivo training models

Several ex-vivo training models are available for the LRYGB. These include bench top models, virtual reality simulators, live animal models and human cadavers. In a survey of 16 Canadian Academic Departments of Surgery and Divisions of General Surgery, the top 5 most effective methods of training in minimally invasive surgery were reported to be basic laparoscopic operations, laparoscopic simulators, animal labs, “black boxes” and advanced laparoscopic operations. Didactic lectures were not perceived to be useful (Chan, Martel, Poulin, Mamazza, & Boushey, 2010).

Bench top models and VR simulators have been well studied in the acquisition of basic laparoscopic skills (Castellvi et al., 2009; Essani et al., 2009; Gonzalez et al., 2010); however, their application for training in LRYGB remains limited (R. Aggarwal, Boza, et al., 2007). One example of a valid bench top model is a fluid-filled porcine small bowel placed in a laparoscopic “black box” to train the creation of a laparoscopic jejunojejunostomy (R. Aggarwal, Boza, et al., 2007).

Anesthetized animals and human cadavers are other training models for advanced laparoscopic skills training. Anesthetized animals offer a unique possibility of teaching advanced laparoscopic skills under in-vivo conditions. However, these animal models are expensive, their anatomical findings differ from human anatomy and animals need to be put down following training. There are obvious societal and ethical concerns regarding training on live animals. In United Kingdom, training on live animal models is not permitted. To overcome some of the shortcomings of an anesthetized animal model, Thiel human cadavers have been proposed for teaching advanced laparoscopic skills (Giger et al., 2008). Thiel preservation offers much better
organ tissue flexibility and plasticity than other preservations. Thiel human cadaver models have been used for training courses in colon, hernia and bariatric surgery with favorable results (Giger et al., 2008). Participants in bariatric surgery courses felt that the Thiel human cadaver model was superior to both the anesthetized porcine model and the VR simulator, because the Thiel human cadaver model offered high operative tactility with tissue color, organ color and consistency comparable to in-vivo conditions (Giger et al., 2008). The Thiel human cadaver models also allow for practice of patient positioning, preparing equipment and establishing a pneumoperitoneum, which cannot be learned on VR simulators. Limitations of the Thiel human cadaver model include the lack of cadavers, the complex and expensive embalming process, and the absence of studies confirming its predictive validity. Consequently, Thiel human cadaver models should only be used for the training in advanced laparoscopic techniques.

Surgical simulation is expensive, but so is training in the operating room. The cost of teaching a laparoscopic, 2 layer enteroenterostomy as part of a LRYGB in the operating room was calculated by using the difference in time for completion of anastomosis between a resident and a staff physician, and amounted to $1,457 USD (Harrington et al., 2007). The cost of teaching 2 anastomoses per year to 15 residents in a residency program was calculated to be $45,061 USD. Additional research is required to address the topic of the cost-effectiveness of surgical simulation in comparison to conventional training.

1.12.4.2.2 Workshops and mini-fellowships

Practicing general surgeons who wish to learn the LRYGB may do so in part by attending workshops and/or mini-fellowships. Both of these training strategies have been proposed as a solution to the problem of increased demand for laparoscopic bariatric surgery and the concurrent lack of laparoscopic bariatric surgeons to perform these operations (Santry, Gillen, & Lauderdale, 2005; T. E. Williams, Jr. & Ellison, 2008).

The impact of eighteen, 2-day laparoscopic bariatric workshops for 300 practicing bariatric and non-bariatric surgeons was assessed by Lord et al (Lord et al., 2006). Each workshop included 7.5 hours of didactic lectures, ½ hour of operating room familiarization, 3 hours of porcine animal lab, 4 hours of viewing live surgery and 1 hour for questions. On a follow-up survey, 45% of participating bariatric surgeons changed from an open to a laparoscopic technique, and 31% of participating non-bariatric surgeons started laparoscopic bariatric surgery. However, 70% of participants felt that additional preceptorship would be required prior to independent
operating. The most useful components of the workshop were surgical demonstrations, the use of new instruments, and the identification and treatment of complications (Lord et al., 2006). Another workshop and a 4-day mini-training program in laparoscopic sleeve gastrectomy showed again that identification and treatment of complications, use of new instruments, as well as surgical and online teaching demonstrations were the most useful aspects (Leandros et al., 2010).

Cottam et al. reported on a 6-week, clinical, laparoscopic bariatric surgery mini-fellowship with a major focus on the LRYGB conducted for board-certified/board-eligible practicing general surgeons with no prior hands-on or formal training in laparoscopic bariatric surgery (Cottam et al., 2007). This mini-fellowship integrated participation in the operating room using a step-wise training approach with outpatient clinics and educational seminars. By the end of the fellowship, each fellow reported a mean of 42 (29-66) cases with the majority being LRYGB. After a mean of 10 months post fellowship completion, 70% of the participants were performing laparoscopic bariatric surgery as part of their practice. The total cost of the mini-fellowship per fellow was $44,695 USD. The breakdown of the total cost ($46,695) per fellow was as follows: state medical licensure ($395), hospital privileges application fee ($300), full malpractice coverage ($8,000), room and board costs (estimated $3,000), fellowship fee ($35,000).

Currently, with an increase in the number of formal laparoscopic bariatric surgery fellowships and the liability concerns regarding workshops and mini-fellowships, the demand for these training modalities seems to be decreasing. The possible explanations for this decrease in demand may include: financial and administrative burdens on the surgeon (Cottam et al., 2007), requirement for additional preceptorship and complementary experience (Clements et al., 2011), and the understanding that a formal workshop or mini-fellowship alone does not provide sufficient training to begin performing laparoscopic bariatric procedures independently (Clements et al., 2011).

The long learning curve for laparoscopic bariatric surgery mandates an urgent search for methods to shorten this curve in the operating room (Christou & Efthimiou, 2009). One method – a framework for post-residency training in advanced laparoscopic surgery - includes four distinct training phases: acquisition of advanced laparoscopic skills, cognitive phase, associative phase and autonomous phase (Lord et al., 2006). Workshops may provide knowledge for the cognitive phase, while mini-fellowships may provide technical skills for the associative phase.
Acquisition of advanced laparoscopic skills and the attainment of necessary procedure-specific cognitive knowledge for laparoscopic bariatric surgery should be done in the form of a standardized, evidence-based ex-vivo technical skills training curriculum.
1.12.4.3 Proposal for an ex-vivo laparoscopic bariatric surgery training curriculum

Despite evidence that ex-vivo technical skills training results in improved performance in the operating room (Grantcharov et al., 2004b; Seymour et al., 2002; Sroka et al., 2010), there is still no formal, evidence-based, ex-vivo, technical skills training curriculum for laparoscopic bariatric surgery. An ideal simulation-based training curriculum should be valid, efficient and proficiency-based (R. Aggarwal, Grantcharov, & Darzi, 2007a). It should systematically take the trainee from the early stages of acquiring cognitive knowledge and basic laparoscopic skills to obtaining privileges for training in the operating room. Our proposal for such a training curriculum and its stages is shown in Figure 12.

Figure 12. Framework for the design of a comprehensive ex-vivo training curriculum in laparoscopic bariatric surgery.

The first stage of the curriculum is Knowledge-Based Learning. Here the trainee acquires procedure-specific knowledge on patient selection, pre-procedure assessment, preparation for the operation, and post-operative management, as well as recognition and management of
complications. Anatomic variations, likely technical errors, ergonomics, safety and limitations of the instruments are also covered. This stage may consist of expert-led didactic sessions, videos and workshops. Evaluation of trainee’s cognitive knowledge can be done in the form of a paper-based and/or oral examination.

The second stage of the curriculum is Deconstruction of the Procedure into Tasks. A hierarchical task analysis can be performed to break down complex laparoscopic procedures (such as LRYGB) into its component tasks and sub-tasks (Sarker, Chang, Albrani, & Vincent, 2008a). Hierarchical task analyses have been created for cholecystectomy (Cao et al., 1999), inguinal hernia repair (Cao et al., 1999), fundoplication (Cao et al., 1999) and laparoscopic colectomy (Sarker et al., 2010). Following the identification of component tasks and sub-tasks, the most challenging and critical components of the procedure can be identified by comparing the performance of novices and experts on each task and sub-task. Procedure-specific assessment tools can then be designed to provide feedback on performance and identify areas of weakness.

The third stage is Training in the Laboratory Environment. The goal is to provide the trainee with the necessary technical skills outside of the operating room to allow for higher level learning (i.e. decision making, progress through the operation, leadership, surgical judgment, etc.) inside the operating room. A combination of bench top models (i.e. cadaveric porcine jejunojejunostomy model (R. Aggarwal, Boza, et al., 2007)), VR simulators and human cadaveric models (i.e. Thiel embalmed cadavers (Giger et al., 2008)) with demonstrated construct validity may be used. Proficiency benchmarks should be defined for each model. A distributed practice schedule with summative feedback should be followed until proficiency is achieved (Moulton et al., 2006). Crisis management and team-based training should also be incorporated at this stage.

The pre-final stage is Transfer of Skills to the Real Environment, where the trainee, who passed the laboratory training component, participates actively in the operating room. In order for this curriculum to be useful and valid, one should demonstrate improved technical performance in the operating room following simulation-based training by comparing individuals who have completed the laboratory training component to those who did not (Grantcharov et al., 2004b). Demonstration of improved patients’ clinical outcomes after implementation of the curriculum would constitute the ultimate proof of validity. Thus, patients’ clinical outcomes should be
tracked before and after the introduction of a simulation-based training curriculum. Trainee’s progress can assessed using the tools designed in stage two.

Demonstration of improvement in clinical patient outcomes after implementation of the curriculum may be quite challenging as the supervisor or proctor will likely prevent most errors and complications in the operating room. For example, if a trainee has placed a poorly positioned suture, the proctor will have it replaced and thereby “prevent” a complication in the outcome. Many poor outcomes will therefore be “prevented” by the proctor. Tracking of intra-operative technical errors, which were corrected by the proctor, is important, as improvements in clinical patient outcomes for trainees under direct supervision may not be detected.

The last stage of the curriculum is Granting Privileges for Operating Room Practice. At this stage, the pass/fail cut-off scores for assessment tools must be defined. The approach for setting cut-off scores can be modeled on the methodology of studies on the Fundamentals of Laparoscopic Surgery curriculum (Fraser et al., 2003; Ritter & Scott, 2007).

If simulation-based training in LRYGBP and laparoscopic bariatric surgery is to be improved, future research studies must be complementary and evidence-based (R. Aggarwal, Grantcharov, et al., 2007a). They should focus on the design and validation of training tools – based on a consensus of international experts – for technical and non-technical skills. These tools can then be incorporated into a comprehensive and valid curriculum focused on training and assessment of all (cognitive, technical and non-technical) components of surgical competency. Implementation of such a curriculum into clinical practice will not only ensure trainee’s proficiency prior to operating room training, but also may shorten learning curves and increase patient safety.
The Delphi method is a widely used and accepted technique for gathering data and building consensus using a series of questionnaires from respondents within their domain of expertise (Hsu & Sandford, 2007). It was developed and used at the Rand Corporation in the 1950’s as a method for achieving convergence of opinion solicited from experts concerning real-world knowledge within specific topic areas. Over the years, this method has been used in various disciplines including medicine, political science, and education (Hsu & Sandford, 2007). The Delphi method can be used to achieve the following objectives (Delbecq, Van de Ven, & Gustafson, 1975):

1) To determine or develop a range of possible program alternatives;
2) To explore or expose underlying assumptions or information leading to different judgments;
3) To seek out information which may generate a consensus on the part of the respondent group;
4) To correlate informed judgments on a topic spanning a wide range of disciplines;
5) To educate the respondent group as to the diverse and interrelated aspects on the topic.

The Delphi method utilizes multiple iterations of a questionnaire concerning a specific topic administered to a select group of opinion leaders in order to achieve consensus among group members. The Delphi method has a number of advantages over other data gathering and analysis techniques including:
1) The iterative and controlled feedback process that allows and encourages the Delphi panel participants to reassess their initial judgments about information provided in previous iterations;
2) The provision of anonymity to the Delphi panel members, thereby reducing the effects of dominant individuals, and minimizing coercion to conform or adopt a certain viewpoint;
3) The suitability of a variety of statistical analysis techniques to interpret the data, thereby allowing for an objective and impartial analysis and summarization of the collected data.

1.13.2 Selection of the Delphi Panel

Selecting the appropriate subjects for the Delphi panel is one of the most important steps within the Delphi method because it directly relates to the quality of the results generated (Judd, 1972). The subjects for the Delphi panel are often selected based upon their area of expertise. There is no exact criterion in the literature concerning the selection of Delphi panel members (Hsu & Sandford, 2007). Individuals are considered eligible to participate in the Delphi panel if they have related backgrounds and experiences concerning the topic at hand. They should be willing to contribute helpful inputs and should be willing to revise their judgments for the purpose of reaching consensus. Hsu and Sandford have suggested that Delphi panel members should be highly trained and competent within the specialized area of knowledge related to the topic at hand. Panel members may be selected based on their publication record and their positions within relevant societies.

The optimal number of subjects to be involved in the Delphi panel is not well delineated in the literature (Hsu & Sandford, 2007). However, if the number of subjects in the Delphi panel is too small, they may not provide a representative pooling of judgments regarding the topic at hand. If the number of subjects in too large, a low response rate may result, and the research team will need to spend greater amount of time analyzing the results from the panel.

1.13.3 Questionnaire Administration

The Delphi method has been designed to permit for multiple iterations of a questionnaire until consensus had been achieved (Hsu & Sandford, 2007). In most cases, three iterations of a questionnaire are often sufficient to collect the required information from the Delphi panel and to reach consensus. The classical Delphi method traditionally begins with an open-ended questionnaire administered in Round 1. The open-ended questionnaire serves to solicit specific
information about a content area from the Delphi panel. Once responses have been received, investigators in charge of the Delphi need to convert the obtained information into a well-structured questionnaire, which in turn is used for Round 2 of data collection. The modified Delphi method employs a structured questionnaire in Round 1 that is based upon an extensive literature review, thereby negating the need for an open-ended questionnaire (Hsu & Sandford, 2007). In Round 2, each Delphi panel member receives a second questionnaire and is asked to rate or rank items to establish preliminary priorities among the items. The responses of the panel members are then analyzed by the research team, and presented back to the panel in Round 3 in the form of a new questionnaire that includes the summary of items and ratings from Round 2. The Delphi panel members are then asked to revise their judgments based on the ratings from Round 2. An additional round of questionnaire administration may be required if consensus was not achieved after Round 3.

1.13.4 Determination of Consensus and Data Analysis

A priori decision criteria must be established to determine consensus among the Delphi subjects. The kind and type of criteria to determine consensus is quite variable in the literature (Hsu & Sandford, 2007). One approach initially ranks the items on the Delphi questionnaire based on their mean rank for that round of the Delphi, and then selects only those items that fall in the highest tertile of ranks (Nathens et al., 2003). Another approach is to select items that greater than 80% of Delphi subjects ranked as 4 or 5 on an Likert scale from 1 to 5 (Palter, MacRae, & Grantcharov, 2011).

The data analysis can involve both qualitative and quantitative approaches. In classical Delphi studies qualitative data is generated from responses to open-ended questions in Round 1, and quantitative data is generated in consecutive rounds. The statistical approach used in Delphi studies examines measures of central tendency (mean, median, and mode) and level of dispersion (standard deviation and inter-quartile range) for questionnaire items, as well as internal consistency among the Delphi subjects (Hsu & Sandford, 2007). Cronbach’s alpha (α) is one statistical index that measures the degree of homogeneity or consistency of opinion among the Delphi subjects. Cronbach’s α is used to quantify the reliability of Delphi subjects. Where the responses of the subjects are highly correlated, they are considered to be internally consistent or homogeneous (Graham, Regehr, & Wright, 2003). Cronbach’s α will be closer to 1.0 as the variance between Delphi subjects becomes smaller in comparison with the variance within each
Delphi subject. Cronbach’s $\alpha$ has been used by several groups for measurement of consensus among Delphi subjects (Graham et al., 2003; Palter, MacRae, et al., 2011).

1.13.5 Weaknesses of the Delphi Method

One weakness of the Delphi method is the potential for a low response rate. This rate is often a result of multiple iterations for the Delphi process, and the time commitment required from the Delphi subjects. With each round of the Delphi process, there is a potential for subject dropout. If a certain proportion of subjects discontinue their responses, the quality of the obtained information could be compromised. Therefore, strategies to address subject motivation are critical to a successful implementation of the Delphi process (Hsu & Sandford, 2007).

Another weakness of the Delphi method is the potential for study investigators to mold the opinions of the Delphi subjects (Hsu & Sandford, 2007). In a modified Delphi method, study investigators have the ability to select what information is presented to the Delphi subjects in Round 1 of the Delphi process. Furthermore, study investigators may also mold the opinions of the Delphi subjects by distorting the feedback that subjects receive after each consecutive round of the Delphi process. In their study, Cyphert and Gant illustrated that Delphi subjects may rank an item on a questionnaire above average after being provided false feedback by study investigators (Cyphert & Gant, 1971). The authors concluded that the Delphi method could be used to mold the opinions of the Delphi subjects.
1.14 Aims and Objectives

The objective of this project was to design and validate a comprehensive, simulation-enhanced training curriculum for a complex minimally invasive operation. This objective was accomplished using three specific aims:

**Aim 1:** To develop a framework for design, validation, and implementation of simulation-based training curricula in surgery using Delphi methodology and consensus of international experts in surgical and medical education.

**Aim 2:**

a) To develop a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery using a hierarchical task analysis, Delphi methodology and international consensus of experienced bariatric surgeons.

b) To demonstrate reliability, validity and feasibility of use for this scale.

**Aim 3:**

a) To design a comprehensive, simulation-enhanced training curriculum for a complex minimally invasive operation addressing cognitive knowledge, technical skills and non-technical skills.

b) To comparing the effect of training in this curriculum on clinical knowledge, technical skills in the operating room and non-technical skill in a simulated operating room environment using a prospective, single-blinded, randomized controlled trial of curriculum trained and conventionally trained mid-level surgery residents.
Chapter 2

A Consensus-Based Framework for Design, Validation and Implementation of Simulation-Based Training Curricula in Surgery

This chapter is modified from the following:

2  A Consensus-Based Framework for Design, Validation and Implementation of Simulation-Based Training Curricula in Surgery

2.1  Abstract

**Background:** Simulation-based training can improve technical and non-technical skills in surgery. To date, there is no consensus on the principles for design, validation and implementation of a simulation-based surgical training curriculum. The aim of this study was to define such principles and formulate them into an interoperable framework using international expert consensus based on the Delphi method.

**Methods:** Literature was reviewed, four international experts in the fields of surgical education and advanced laparoscopic surgery were queried, and face-to-face consensus conference of national and international members of surgical societies was held to identify the items for the Delphi survey. Forty-five international experts in surgical education were invited to complete the online survey by ranking each item on a Likert scale from 1 to 5. Consensus was predefined as Cronbach's alpha >= 0.80. Items that 80% of experts ranked as >= 4 were included in the final framework.

**Results:** Twenty-four international experts with training in general surgery (11), orthopaedic surgery (2), obstetrics and gynaecology (3), urology (1), plastic surgery (1), pediatric surgery (1), otolaryngology (1), vascular surgery (1), military (1) and doctorate level educators (2) completed the iterative online Delphi survey. Consensus among participants was achieved after one round of the survey (Cronbach’s alpha = 0.91). The final framework included pre-development analysis; cognitive, psychomotor and team-based training; curriculum validation evaluation and improvement; and maintenance of training.

**Conclusions:** The Delphi methodology allowed for determination of international expert consensus regarding the principles for design, validation and implementation of a simulation-based surgery training curriculum. These principles were formulated into a framework, which can be used internationally across surgical specialties as a step-by-step guide for the development and validation of future simulation-based training curricula.
2.2 Introduction

The number of research studies addressing the role of simulation in surgical education has grown dramatically over the last decade. Multiple authors have reported on the beneficial effects of simulation-based training for technical (Ahlberg et al., 2007a; Fried, 2006; Korndorffer et al., 2005; Seymour, 2008) and non-technical (Black et al., 2010; Kozmenko, Paige, & Chauvin, 2008; Merry, 2007) skills. Two systematic reviews addressing this topic have also been published (Issenberg et al., 2005; Sutherland et al., 2006). However, despite the evidence from randomized controlled trials for the use of simulation in surgical education (Grantcharov et al., 2004a; Larsen et al., 2009; Seymour et al., 2002), most surgery training programs continue to struggle with the integration of structured, proficiency-based training into their curricula (Korndorffer, Stefanidis, & Scott, 2006; Stefanidis & Heniford, 2009). These difficulties are in part a result of the multitude of opinions on what should be included in such curricula, how they should be validated and how they should be integrated into the residency training programs. The result has been a multitude of different curricula (frequently on the same procedures) being developed with competing, contradictory and duplicative approaches. The Association of Program Directors in Surgery and the American College of Surgeons have jointly tackled this issue by developing the ACS/APDS Surgical Skills Curriculum for Residents, which includes 20 modules that address basic surgical skills, 15 modules that address advanced surgical skills and procedures, and 10 modules that address team-based skills ("ACS/APDS Surgical Skills Curriculum for Residents," 2010).

Several authors have emphasized the importance of teaching cognitive knowledge in a comprehensive curriculum for training of surgical skills (Satava, 2005; Tang, Hanna, & Cuschieri, 2005; Van Herzeele et al., 2008). Such teaching should focus on the tasks required for correct execution of a particular procedure, including error detection, forward planning and decision making. Tang and colleagues demonstrated that the majority of errors performed by trainees during simulated training were not caused by a technical mistake, but rather by a knowledge gap in understanding of the correct sequence of steps in the particular task (Tang et al., 2005). A non-randomized study of the impact of cognitive training on task execution emphasized that cognitive training not only improved understanding of a particular task, but also improved its execution (Van Herzeele et al., 2008).

A comprehensive, simulation-based training curriculum should incorporate evidence-based methodological principles, such as deliberate practice, distributed practice schedule and
proficiency-based training. Participation in such a curriculum should also be made mandatory (Chang, Petros, Hess, Rotondi, & Babineau, 2007; Stefanidis, Acker, Swiderski, Heniford, & Greene, 2008). K. Anders Ericsson has studied the concept of deliberate practice and its impact on the acquisition of expert performance (Ericsson, 2004, 2008). He suggested five conditions which are required for deliberate practice: (1) a well-defined task to practice; (2) provision of detailed and immediate feedback; (3) motivation to improve; (4) variable level of difficulty for the chosen task; and (5) ample opportunity for repetition and gradual refinement of performance. Moulton and colleagues studied massed versus distributed practice in a randomized controlled trial of 38 surgery residents learning a new skill of microvascular anastomosis (Moulton et al., 2006). Residents who practiced in a distributed pattern exhibited superior retention of technical skills. Proficiency-based training has been shown to result in an improved operative performance (Seymour, 2008; Seymour et al., 2002) and a decrease in error rate in the operating room (Ahlberg et al., 2007a).

Aggarwal and colleagues, as well as Stefanidis and Heniford have tried to incorporate some of the above-mentioned methodological principles into frameworks for the design of simulation-based training curricula in surgery (R. Aggarwal, Grantcharov, & Darzi, 2007b; Stefanidis & Heniford, 2009). Aggarwal’s 5-step framework includes a step-wise progression from knowledge-based learning, to deconstruction of the procedure into its component tasks, to training in a laboratory environment, to demonstrating transfer of skills to the real environment, and to eventually granting privileges for independent practice (Figure 13). The framework of Stefanidis and Heniford takes a trainee from baseline assessment of technical skill, to demonstration of procedure using video tutorials, to deliberate practice on a simulator in distributed sessions with summary feedback, and finally to attainment of proficiency with overtraining, maintenance of training and post-training assessment (Figure 14). Both frameworks are similar in their recommendations for combining cognitive learning with training to proficiency on simulated models using deliberate practice and a distributed practice schedule.
The purpose of this study was to advance the field of surgical education and curricular development by developing a standardized, prescriptive and internationally relevant framework for the design, validation and implementation of a simulation-based surgical training curriculum. It used current evidence-based methodological principles for simulation-based training and international expert consensus via the Delphi method.
2.3 Methods

2.3.1 Study Design

Classical and modified Delphi methods were used to achieve consensus among a panel of international experts in surgical simulation on the essential components of a standardized and comprehensive framework for the design, validation and implementation of a simulation-based surgical training curriculum. The classical Delphi method - developed by the RAND Corporation in 1948 – provides a means for obtaining opinions of experts in a systematic manner (Fink, Kosecoff, Chassin, & Brook, 1984; Nathens et al., 2003; P. L. Williams & Webb, 1994). It is an anonymous process where ideas are presented to participants in the form of a questionnaire (Graham et al., 2003). The responses are collated, analyzed and presented back to the participants in an iterative fashion until consensus is achieved (Graham et al., 2003). The modified Delphi method has the same basis; however, it is conducted in a planned consensus conference with participation of experts in the subject using the same iterative methodology.

2.3.2 Participant Selection

Participants were identified as opinion leaders in the field of surgical education as evidenced by their key roles within North American, South American, European, Asian, and Australasian surgical societies and organizations (Table 3). All participants were members of the Alliance for Surgical Simulation Education and Training (ASSET) group – a group of key members of senior leadership from a diverse cross-section of 16 surgery societies in the United States and 9 internationally. Accrediting organizations, including US Department of Defense and Veterans Health Administration were also represented. These individuals were purposefully invited for panel membership to represent opinions from different surgery specialties across a diverse geographic area (Table 4), though their opinions were not the official opinions nor endorsement of their respective societies. To ensure adequate representation of various surgery specialties across a wide geographic area, 45 individuals from 8 different surgical and 3 non-surgical specialties from 9 countries were invited. Participation was voluntary, and informed consent was implied if a panel member chose to participate.
Table 3. Surgery societies and organizations surveyed.

<table>
<thead>
<tr>
<th>Location</th>
<th>Society and organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>American Academy of Orthopaedic Surgeons</td>
</tr>
<tr>
<td></td>
<td>American Academy of Otolaryngology - Head and Neck Surgery</td>
</tr>
<tr>
<td></td>
<td>American Association for Thoracic Surgery</td>
</tr>
<tr>
<td></td>
<td>American Association of Gynecologic Laparoscopists</td>
</tr>
<tr>
<td></td>
<td>American Association of Plastic Surgeons</td>
</tr>
<tr>
<td></td>
<td>American College of Obstetricians and Gynecologists</td>
</tr>
<tr>
<td></td>
<td>American College of Surgeons</td>
</tr>
<tr>
<td></td>
<td>American Hernia Society</td>
</tr>
<tr>
<td></td>
<td>American Society of Colon and Rectal Surgeons</td>
</tr>
<tr>
<td></td>
<td>American Society of Plastic Surgeons</td>
</tr>
<tr>
<td></td>
<td>American Urogynecologic Society</td>
</tr>
<tr>
<td></td>
<td>American Urological Association</td>
</tr>
<tr>
<td></td>
<td>Arthroscopy Association of North America</td>
</tr>
<tr>
<td></td>
<td>Association for Surgical Education</td>
</tr>
<tr>
<td></td>
<td>Society of American Gastrointestinal and Endoscopic Surgeons</td>
</tr>
<tr>
<td></td>
<td>Center for Advanced Medical Learning and Simulation</td>
</tr>
<tr>
<td></td>
<td>Society of Laparoscopic Surgeons</td>
</tr>
<tr>
<td></td>
<td>US States Department of Defense</td>
</tr>
<tr>
<td></td>
<td>Veterans Health Administration</td>
</tr>
<tr>
<td></td>
<td>Thoracic Surgery Directors Association</td>
</tr>
<tr>
<td></td>
<td>National SimLEARN Center of the Veterans Health Administration</td>
</tr>
<tr>
<td>South America</td>
<td>Latin American Association of Laparoscopic Surgery</td>
</tr>
<tr>
<td></td>
<td>Latin American Hernia Foundation</td>
</tr>
<tr>
<td>Europe</td>
<td>Royal College of Surgeons of England</td>
</tr>
<tr>
<td></td>
<td>Royal College of Surgeons in Ireland</td>
</tr>
<tr>
<td>Asia</td>
<td>Asia Pacific Hernia Society</td>
</tr>
<tr>
<td>Australasia</td>
<td>China Hernia Society</td>
</tr>
<tr>
<td>Africa</td>
<td>Royal Australasian College of Surgeons</td>
</tr>
<tr>
<td></td>
<td>South African Society for Obstetricians and Gynaecologists</td>
</tr>
</tbody>
</table>
Table 4. Composition of the Delphi international expert panel.

<table>
<thead>
<tr>
<th>Location</th>
<th>Surgical and nonsurgical specialty</th>
<th>No. of participants contacted</th>
<th>No. of participants responding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>General surgery</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>USA</td>
<td>General surgery</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Obstetrics and gynecology</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Orthopaedic surgery</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Urology</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Plastic surgery</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Military</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>PhD, MPH</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Otolaryngology</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Internal medicine</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>UK</td>
<td>General surgery</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Brazil</td>
<td>General surgery</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Australia</td>
<td>General surgery</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ireland</td>
<td>General surgery</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Vascular surgery</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>China</td>
<td>General surgery</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>South Africa</td>
<td>Obstetrics and gynecology</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>General surgery</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Pediatric surgery</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
2.3.3 Derivation of Survey Items

Survey items were based on evidence-based educational principles, published peer-reviewed literature, results of a modified Delphi consensus conference of the ASSET group in June of 2011 and opinions of four international experts in surgical education. A complete list of survey items is presented in Table 5.

2.3.4 Survey Administration

A complete list of survey items (Table 5) was presented to each participant after the 1st consensus conference in the form of an online survey (www.SurveyMonkey.com). Each participant was asked to rate each item on a Likert-type scale from 1 (“strongly disagree”) to 5 (“strongly agree”) with respect to the degree that s/he believed that item should be included in the final framework. Participants were given the opportunity to comment on each item and to clarify their ratings. One e-mail reminder to complete the survey was sent 2 weeks after the initial invitation e-mail. Data collection stopped 31 days after the original invitation e-mail.

2.3.5 Determination of Consensus

A classical Delphi methodology analysis was used to establish consensus among the participants (Graham et al., 2003; Palter, MacRae, et al., 2011). On the basis of the work of Graham et al.(Graham et al., 2003) and Palter et al.(Palter, MacRae, et al., 2011), Cronbach’s alpha was chosen as the statistical index to quantify a measure of consensus (Bland & Altman, 1997). Cronbach’s alpha of 0.80 was chosen to represent an acceptable measure of consensus for this study.

2.3.6 Creation of Final Framework

Once consensus among study participants was achieved (Cronbach’s alpha >/= 0.80), items that were rated as 4 (“agree”) or 5 (“strongly agree”) by greater than 80% of participants were selected for inclusion into the final framework. Comments for each survey item were reviewed and modifications to the items were made based on the comments. The items in the final framework were then discussed and revised during a second modified face-to-face Delphi consensus conference of ASSET in December 2011.
Table 5. Complete list of Delphi survey items.

**Pre-Development Analysis**
1. Perform a Needs Assessment / Define the Problem
   a. Determine the goals and objectives for the curriculum
   b. Determine the desired learning outcomes
   c. Determine the desired competencies to be achieved
   d. Determine the population of interest (novice, intermediate, or advanced trainees, or practicing surgeons)
2. Determine the available resources at your institution (within different departments)
   a. Knowledge content
   b. Physical space
   c. Equipment (existence of simulators? Type? High or low fidelity?)
   d. Available personnel
   e. Financial resources
   f. Support of faculty members

**Curriculum Development**

**Cognitive Component**
1. Pre-test of procedure-specific knowledge
   a. Written test format
      i. Multiple choice
      ii. Short answer questions
      iii. Essay type questions
   b. Oral examination (case based)
   c. Other
2. Pre-procedure patient assessment and preparation
   a. Review of indications for surgery
   b. Patient consent
   c. Teamwork and environment considerations
3. Pathobiology of disease
4. Procedural-specific operative knowledge
   a. Key steps of the procedure
   b. Anatomic variant knowledge
      i. Descriptions of tissue planes
      ii. Important structures
   c. Instruments
      i. Types
      ii. Indications and Safety
      iii. Limitations
   d. Ergonomics
   e. Demonstration of fundamental technical skills needed to complete the procedure
   f. Operative management
      i. Anticipations of errors
      ii. Recognition of errors
      iii. Strategies for management of errors
      iv. Inclusion of video examples of failures / disasters
   g. Post-operative management
      i. Knowledge of potential complication
      ii. Recognition of complications
      iii. Management of complications
5. Time for questions to the expert, answers, and reflections by trainees
6. Post-test of learned procedure-specific knowledge
   a. Written test format
      i. Multiple choice
ii. Short answer questions  
iii. Essay type questions  

b. Oral examination (case based)  
c. Other  

7. Delivery of the cognitive component  
   a. Didactic sessions  
      i. Faculty-led  
      ii. Trainee-led  
   b. Procedure-specific video-based tutorials  
      i. Pauses in the video at critical steps of the procedure  
      ii. Faculty led  
      iii. Self-directed  
   c. Self-directed readings  
      i. Selected journal articles  
      ii. Selected text-book chapters  

Technical Component  

1. Deconstruction of the procedure of interest into its component tasks/steps  
   a. Acquire video-recordings of experts performing the procedure of interest  
      i. Include most common approaches  
      ii. Include all possible approaches  
   b. Perform a hierarchical tasks analysis (HTA) of the procedure  
      i. Two or more independent reviewers to watch the video-recordings and write down all the steps and sub-steps of the procedure  
      ii. Establish consensus between reviewers on the steps and sub-steps  
      iii. Each task should have a clear and unambiguous start point and end point  
      iv. By reviewing the video-recordings of novices, identify which tasks are the most difficult for a novice to perform  

2. Create a procedure-specific assessment tool using the identified steps and sub-steps  
   a. Use Delphi methodology to identify the steps and sub-steps from the HTA to be included in the technical skills assessment tool  
   b. Establish reliability of the tool  
   c. Establish validity of the tool  
      i. Demonstrate construct validity by demonstrating a difference in scores of novice surgeons (< 10 procedures), intermediates (10-100 procedures), and experts (>100 procedures)  
   d. Define a cut-off value to be used in summative assessment to define competency  

3. Conduct initial assessment of trainees technical skill using the procedure-specific assessment tool  

4. Identify existing simulation models for training in the key steps of the procedure (Reznick & MacRae, 2006)  
   a. Basic skills training  
      i. Synthetic models  
      ii. Computer based models  
         1. low fidelity  
         2. high fidelity  
      iii. Cadaveric models  
      iv. Live Animal models  
   b. Advanced (procedure-specific) skills training  
      i. Synthetic models  
      ii. Computer based models  
      iii. Cadaveric models  
      iv. Live animal models  
   c. Confirm or define validity for the chosen models  
      i. Face  
      ii. Content
iii. Concurrent
iv. Construct
v. Predictive

5. Define different levels of difficulty for each simulation model

6. Establish expert proficiency benchmarks for each simulation model
   a. Using measurement parameters imbedded in the model (i.e. virtual reality simulators)
   b. Using procedure-specific assessment tool developed previously

7. Set expert proficiency benchmarks as the “passing” score for the curriculum criterion for the curriculum
   a. Forced proficiency at each level before progression to the next level (learn basic skills before moving on to advanced levels)

8. Define a specific practice schedule for trainees
   a. Distributed practice
   b. Deliberate practice
   c. Variability of practice (variable levels of difficulty)
   d. Practice until expert level of proficiency is achieved

9. Provide extrinsic feedback (during practice sessions) to trainees on their performance (Sarker & Patel, 2007)
   a. Review of performance on video
   b. Review of video-recording demonstrating expert level of performance
   c. Summary expert feedback
   d. Review areas of weakness based on the procedure-specific assessment tool

10. Include a component of overtraining after reaching proficiency (Stefanidis & Heniford, 2009)
    a. Document achievement of expert level proficiency on 2 consecutive practice sessions
    b. Perform 5-10 additional practice sessions for reinforcement

Team-based Component

1. Develop a module for teaching non-technical skills (Teamwork/Communication) as related to the procedure of interest
   a. Identify the roles and responsibilities of allied health care team
   b. Define which non-technical skills must be learned for the procedure of interest
   c. Develop crisis-based scenarios for learning non-technical skills in a simulated environment
   d. Identify an assessment tool to be used for evaluation of non-technical skills
      i. Formative assessment
      ii. Summative assessment
   e. Develop specific communication protocols

Curriculum Validation, Evaluation and Improvement

Validation

1. Demonstrate transfer of skills to the real environment
   a. Conduct a randomized controlled trial
      i. Compare trained group with untrained group
      ii. Demonstrate difference in cognitive knowledge and technical skill
      iii. Evaluate the impact of simulation-based training on the learning curve in the operating room
      iv. Calculate the Transfer-Effectiveness Ratio for the training curriculum

Evaluation and Improvement

1. Evaluate the curriculum
   a. Strengths and weaknesses
   b. Areas for improvement
   c. Measure patient outcomes in the operating room (morbidity and mortality) and compare these to pre-curriculum levels
2. Develop methods for continuous program improvement
**Granting Privileges for Independent Practice**

1. Define cut-off values for the cognitive test and a procedure-specific rating scale to be used in summative assessment
2. Define cut-off values on the assessment scale for non-technical skills to be used in summative assessment

**Maintenance Training**

1. Test cognitive knowledge and technical skills at xxx months intervals following the completion of curriculum training to define the time period at which knowledge and technical skills begin to deteriorate
2. Define a regular testing/training schedule to ensure maintenance of learned knowledge and technical skills
3. Include:
   a. Case presentations (increasing in complexity over time)
   b. Discussions with experts
   c. Mentoring of junior residents
   d. Continued formative feedback

**2.3.7 Data Analysis**

Descriptive statistics were calculated for all survey items. Cronbach’s alpha was calculated as a measure of consensus among study participants. Missing data points, which resulted from submissions of incomplete surveys, were dealt with in three ways: (1) a mean participant score for that item was calculated, and the missing data point was replaced with the mean score; (2) the missing data point was replaced with a 3 (“neutral”) based on the assumption that because the question was omitted, the participant did not feel strongly about including or excluding that item from the final framework; and (3) the missing data point was replaced with the mode for that item. Cronbach’s alpha was calculated for each of the above methods. All statistical analyses were performed using STATA version 12.0 (College Station, Texas, USA).
2.4 Results

Forty-five international experts in surgical education from 9 specialties and 9 countries were invited to participate in this study (Table 4). Twenty-four (53%) experts from 8 specialties and 7 countries completed the online survey (Table 4). The calculated Cronbach’s alpha was 0.91 (missing data points replaced by mean), 0.89 (missing data points replaced by 3 “neutral”), and 0.91 (missing data points replaced by mode). A second round of the Delphi survey was not necessary, because consensus was achieved during the 1st round. The main components of the final framework included: pre-development analysis; cognitive, psychomotor and team-based training; curriculum validation, evaluation and improvement; and maintenance of training (Table 6). Cronbach’s alpha for each individual component of the framework ranged from 0.83 to 0.99 (Table 7). The percentage of missing data for each individual component ranged from 0 to 14 percent (Table 7).
Table 6. An evidence-based framework for the design of a simulation-based surgical training curriculum.

**Pre-Development Analysis**

1. Perform a Needs Assessment / Define the Problem
   a. Determine the goals and objectives for the curriculum
   b. Determine the desired learning outcomes measures and proficiencies to be achieved
      i. Include Governing authorities - Boards, Societies, ACGME etc. when possible
   c. Determine the common errors that need to be taught
   d. Define the population of interest (novice, intermediate, advanced trainee, expert surgeon)
   e. Include Program Director, Faculty, Educator, User (resident, med student, etc.) and Board

2. Determine the available resources at your institution, society, etc (OPTIONAL)
   a. Knowledge content
   b. Physical space
   c. Equipment (Simulators? Types? High or low fidelity?)
   d. Personnel including faculty with formal educator expertise (i.e. fellowship, masters, or PhD)
   e. Financial resources
   f. Support of faculty members, chair, chief, and/or program-director

**Curriculum Development**

**Cognitive Component (focused upon specific technical issues)**

1. Pre-test of procedure-specific knowledge
   a. Written test format
      i. Multiple choice
   b. Interactive, graphic-based image software that delivers procedural training

2. Pre-procedure patient assessment and preparation
   a. Review of indications / contraindications for surgery
   b. Teamwork and environment considerations (functioning equipment, instruments, suture/stapling devices, etc.)

3. Procedural-specific operative knowledge
   a. Key steps of the procedure
   b. Anatomy: Descriptions of important structures and tissue planes
   c. Instruments: Identify types, indications, safety and limitations of instruments
   d. Psychomotor Skills: Demonstration of fundamental technical skills needed to complete the procedure
   e. Errors (most common and most critical though infrequent):
      i. Anticipation / avoidance
      ii. Recognition
      iii. Management
   f. Post-operative Complications (potential and actual):
      i. Knowledge of
      ii. Anticipation / Avoidance
      iii. Recognition and management
g. Inclusion of video examples of failures / disasters

4. Faculty Interaction: Time for questions to the expert; answers and reflections by trainees

5. Post-test of learned procedure-specific knowledge
   a. Written test format
      i. Multiple choice questions
      ii. Interactive, graphic-based image software that delivers procedural training
   b. Mock operative note dictation

6. Delivery of the cognitive component
   a. Procedure-specific video-based tutorials with pauses in the video at critical steps of the procedure
   b. Interactive software that delivers procedural training
      i. Interactive graphic-based image of procedure has to be identical to psychomotor skill simulator (this is critical)
   c. Progress to the psychomotor skills component of the training is not permitted until cognitive part is passed, preferably with 100%.

Psychomotor Skills (Technical Component)

1. Deconstruction of the procedure of interest into its component tasks / steps
   a. Acquire video-recordings of different approaches of experts performing the same procedure of interest
      i. Include only the tasks in common from multiple different approaches
   b. Perform a hierarchical tasks analysis (HTA) of the procedure (ideally should include experts in surgical education, behavioral psychologists, expert clinicians)
      i. Two or more independent reviewers to watch the video-recordings and write down all the tasks / steps of the procedure
      ii. Establish consensus between reviewers on the tasks / steps
      iii. Each tasks / steps should have a clear and unambiguous definition of the task / subtask, start point and end point, and quantitative measure (if possible) or clearly defined Likert Scale
      iv. By reviewing the video-recordings of novices, identify which tasks / steps are the most difficult for a novice to perform and what are the most common errors

2. Create a procedure-specific assessment tool using the identified tasks / steps / errors
   a. Use Delphi methodology to identify the tasks / steps / errors from the HTA to be included in the technical skills assessment tool
   b. Establish reliability of the tool
d. Establish validity of the tool
e. Demonstrate construct validity by demonstrating a difference in scores of novice surgeons (< 10 procedures), intermediates (10-100 procedures), and experts (>100 procedures). The number of procedures at each level may vary significantly depending on the type of procedure.
f. Establish the benchmark criteria for competency (Experts’ mean +/- 1 standard deviation)
   i. Analyze experts’ learning curves (stop when 2 consecutive trials show no improvement – i.e. they have reached a plateau)
ii. Establish the mean of all the experts’ plateaus
iii. Establish the mean +/- 1 standard deviation of the mean of all the experts
iv. The benchmark of competence is experts’ mean minus 1 standard deviation
g. Define a cut-off value to be used in formative assessment (from the errors list)
h. Define a cut-off value to be used in summative assessment to define proficiency (from the benchmark criteria)

3. Conduct initial assessment of trainees technical skill using the procedure-specific assessment tool

4. Identify existing simulation models for training in the key tasks / steps of the procedure
   a. If no procedure-specific simulator exists, determine how to develop the desired simulator
   b. Basic skills training
      i. Synthetic models
      ii. Low fidelity computer based models
      iii. Aim to select models with embedded objective metrics
   c. Advanced (procedure-specific) skills training
      i. Synthetic models
      ii. Computer based models
      iii. Cadaveric / tissue models
      iv. Live animal models
   d. Confirm or define validity for the chosen models
      i. Face
      ii. Content
      iii. Concurrent
      iv. Construct
      v. Predictive

5. Define different levels of difficulty for each simulation model

6. Establish expert proficiency benchmarks for each simulation model (see item 2.e.)
   a. Using measurement parameters imbedded in the model
   b. Using the previously-developed procedure-specific assessment tool

7. Set acceptable level of proficiency as the “pass” score for each model in the curriculum (see item 2.e.)
   a. Forced proficiency at each level before progression to the next level (learn basic skills before move on to advanced levels)

8. Define a specific practice schedule for trainees
   a. Distributed practice
   b. Deliberate practice
   c. Variability of practice (variable levels of difficulty)
   d. Practice until acceptable level of proficiency is achieved
   e. Define amount of overtraining (if needed)

9. Provide extrinsic feedback (during practice sessions) to trainees on their performance
a. Formative
   i. Immediate computer generated feedback
   ii. Immediate, face-to-face feedback by expert instructors with alternative approaches demonstrated and practiced in the presence of the instructor
b. Summative
   i. Review areas of weakness based on the procedure-specific assessment tool
   ii. Review of performance on video

Team-based Component (supports technical aspect of procedure)
1. Develop a module for teaching non-technical skills (teamwork, communication, situation awareness, etc.) as related to the procedure of interest
   a. Identify the roles and responsibilities of allied health care team (individual and team training)
   b. Develop crisis-based scenarios for learning non-technical skills in a simulated environment (appropriate for level of learner and of incremental difficulty)
   c. Use available assessment tools for formative evaluation of non-technical skills (evaluate the reliability, validity and applicability of instruments; if not available develop new tools)
   c. Use an existing team-training model (eg. TeamSTEPPS) if available
d. Incorporate train-the-trainer programs for teaching faculty
   i. Provide specific training for debriefing skills

Curriculum Validation, Evaluation and Improvement

Validation
1. Conduct a randomized controlled trial to demonstrate transfer of learned skills to the real environment
   a. Compare simulator trained group with traditionally trained group
   b. Demonstrate difference in cognitive knowledge and technical skill
   c. Evaluate the impact of simulation-based training on the learning curves in the operating room
   d. Calculate the Transfer-Effectiveness Ratio (TER) for the training curriculum

Evaluation and Improvement
1. Prospectively collect data on curriculum validity
   a. Compare data of pre-post test learner performance
   b. Learner retention data
   c. Learner feedback
   d. Impact of curriculum on clinical outcomes
2. Perform periodic evaluations of curriculum and adjust curriculum based on feedback/experience
   a. Identify strengths and weaknesses and areas for improvement
   b. Develop methods for continuous program improvement

Maintenance of Training
1. Define post training intervals at which tests of cognitive knowledge and technical skills will be administered to assure ongoing proficiency of the learner
2. Develop a specific “retraining” curriculum, based on above principles. Generally, this will be less extensive and more focused than the full curriculum.
3. Include retraining to proficiency during testing sessions (maintenance practice) with formative feedback.
4. Study skill degradation for particular tasks.
Table 7. Agreement among the participants on each component of the framework.

<table>
<thead>
<tr>
<th>Framework component</th>
<th>Cronbach’s α</th>
<th>Missing data, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predevelopment analysis</td>
<td>0.89</td>
<td>0</td>
</tr>
<tr>
<td>Cognitive component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest of procedure-specific knowledge</td>
<td>0.97</td>
<td>0</td>
</tr>
<tr>
<td>Preprocedure patient assessment and preparation</td>
<td>0.89</td>
<td>0</td>
</tr>
<tr>
<td>Procedure-specific operative knowledge</td>
<td>0.83</td>
<td>0</td>
</tr>
<tr>
<td>Post-test of learned procedure-specific knowledge</td>
<td>0.95</td>
<td>0</td>
</tr>
<tr>
<td>Delivery of the cognitive component</td>
<td>0.92</td>
<td>0</td>
</tr>
<tr>
<td>Psychomotor skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deconstruction of the procedure of interest into its component tasks/steps</td>
<td>0.89</td>
<td>12.0</td>
</tr>
<tr>
<td>Create a procedure-specific assessment tool using the identified tasks/steps/errors</td>
<td>0.84</td>
<td>8.0</td>
</tr>
<tr>
<td>Identify existing simulation models for training in the key tasks/steps of the procedure</td>
<td>0.91</td>
<td>12.0</td>
</tr>
<tr>
<td>Establish expert-level proficiency benchmarks for each simulation model</td>
<td>0.99</td>
<td>8.0</td>
</tr>
<tr>
<td>Define a specific practice schedule for trainees</td>
<td>0.90</td>
<td>12.0</td>
</tr>
<tr>
<td>Provide extrinsic feedback (during practice sessions) to trainees on their performance</td>
<td>0.86</td>
<td>12.0</td>
</tr>
<tr>
<td>Team-based component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop a module for teaching nontechnical skills</td>
<td>0.86</td>
<td>8.0</td>
</tr>
<tr>
<td>Curriculum validation, evaluation, and improvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td>0.91</td>
<td>12.0</td>
</tr>
<tr>
<td>Evaluation and improvement</td>
<td>0.99</td>
<td>12.0</td>
</tr>
<tr>
<td>Maintenance of training</td>
<td>0.88</td>
<td>16.0</td>
</tr>
</tbody>
</table>
2.5 Discussion

This study used classical and modified Delphi methodology to establish consensus among a panel of international experts in surgical education on the principles for design, validation and implementation of a simulation-based surgical training curriculum. These principles were then formulated into a generic, prescriptive, uniform and evidence-based framework for international implementation in any surgical specialty and for any surgical procedure of interest. Solicitation of experts from most surgical specialties and doctorate level educators, from civilian and military backgrounds, and from diverse geographical locations (North America, South America, Europe, Australia and New Zealand) strengthened the content validity of the developed framework and included perspectives not often included in curriculum development.

The final framework (Table 6) has some similarities with the curricula development frameworks proposed by Aggarwal and colleagues (R. Aggarwal, Grantcharov, et al., 2007b), as well as Stefanidis and Heniford (Stefanidis & Heniford, 2009). These individual frameworks represent excellent analysis and recommendations from a single perspective of experience and evidence, although they lack a broad international component with input from multiple surgical specialties and regulatory organizations. The proposed framework in this study benefits from these pioneering studies, adds a consensus of 24 international experts from most surgical specialties, and is grounded in evidence-based methodological principles of educational theory. In addition to these fundamental principles, it supports the Dreyfus and Dreyfus model of training from novice to expert (Dreyfus et al., 1986). The comprehensive incorporation of all of these components is a major strength of the proposed framework.

The broad principals of educational theory constitute the infrastructure of this proposed framework. These principals have been previously reviewed in detail elsewhere (Drew et al., 1999; Kneebone, 2003; Kovacs, 1997). They include acquisition of cognitive knowledge (Satava, 2005; Tang et al., 2005; Van Herzeele et al., 2008), psychomotor skills training, and non-technical skills training (communication, collaboration, professionalism, and management). Acquisition of cognitive knowledge can be organized into phases of conceptualization, visualization and verbalization (Kovacs, 1997). During the conceptualization phase, the learner learns the relevant anatomy, indications, contraindications and complications related to the procedure. During the visualization and verbalization phase, the learner is required to describe the procedure from start to finish. Psychomotor skills training requires appropriate targets for proficiency (Ahlberg et al., 2007a), deliberate (Ericsson, 2004, 2008) and distributed practice
Moulton et al., 2006), motivation and timely feedback (Kneebone, 2003), overtraining, and maintenance training (Stefanidis & Heniford, 2009). Communication, collaboration, professionalism, and management are the non-technical components of surgical competency. These components should also be made part of a comprehensive surgical training curriculum.

The classical Delphi method used in this study has several strengths. It offers the opportunity to conduct the questionnaire by mail or online; thereby improving feasibility and lessening costs (Fink et al., 1984). Participants can be recruited from various geographical locations and clinical backgrounds (Graham et al., 2003), whereas the anonymous nature of the Delphi method makes it challenging for a single influential participant to have a disproportionate impact on the survey’s outcome (Graham et al., 2003). The use of classical Delphi method allowed for conclusion of this study at a minimal cost. Participants were contacted over e-mail, and the survey was administered online. A Cronbach’s alpha value of greater than 0.90 in this study implies that there were enough participants to achieve consensus, as well as a lack of controversy in regards to the proposed framework items. This lack of controversy is a result of the effort on the part of the authors to include only those items which were based on valid and evidence-based principles in surgical education.

The present study has a few limitations. The response rate for this study (53%) was moderate, albeit acceptable and consistent with survey response rates in health professional literature (Burns et al., 2008; Palter, MacRae, et al., 2011; Sierles, 2003). It is possible that participants who chose not to participate in the survey were not checking their e-mail regularly, were not comfortable with online surveys or missed the reminder in their inbox. Extending the duration of data collection, incorporating additional e-mail and telephone reminders, and sending out the survey by mail could have potentially increased the response rate. However, to maintain participants’ interest in the survey and anticipating a possible need for a second round of the survey, data collection was stopped after 31 days. With the finding of consensus among panel members after the first round of the survey, a second round was not required. In addition, the response rate must be judged with the knowledge of the population that this study attempted to recruit – international experts with senior leadership positions in various national and international surgical societies.

The distribution of participants’ geographical locations was not uniform. The greatest proportion of experts was from North America, whereas continental Europe, Asia and South America were
underrepresented. Consequently, results likely reflect an Anglo-American point of view and should thus be interpreted in the context of an Anglo-American education format. Similarly, other surgery groups such as ophthalmology, neurosurgery, and oral maxillofacial surgery were not represented in this study. Future studies can address these limitation by surveying ophthalmology, neurosurgery and oral maxillofacial surgery specialties and underrepresented geographical regions – using a Delphi method – and then pooling the results.

The classical Delphi method, although considered to be one of the ideal means to elicit expert opinion and determine consensus, has been criticized because the breadth of the question under consideration is in part controlled by the investigators (Graham et al., 2003; Palter, MacRae, et al., 2011). This study mitigated this criticism by deliberately constructing the Delphi survey to contain not only the preliminary information from the initial ASSET consensus conference, but also items from published, peer-reviewed literature and recommendations from interviews with key informants. The reliability of the results obtained using a Delphi method has also been brought into question (Keeney, Hasson, & McKenna, 2001). However, reports of Delphi results being accurate 16 years after the initial survey administration suggest that results are reliable (Ono & Wedemeyer, 1994). The present study attempted to mitigate this potential limitation of reliability by recruiting participants from a diverse geographical area and from most surgery specialties. Missing data points from partial completion of the survey represent another limitation; however, regardless of how the missing data points were analyzed, Cronbach’s alpha remained well above the predetermined cut-off of 0.80 – suggesting consensus among the participants.

Lastly, it is important to discuss and clarify the role for simulation in surgery residency-training programs. Over the past 100 years, time-based apprentice-type surgery training programs have produced many superb surgeons; however, with restrictions in resident work hours, increasing emphasis on patient safety and increasing costs for training in the operating room, trainees now have fewer opportunities to train in the operating room. Simulation-based training is not intended to completely replace training in the operating room; rather, it is intended to provide an effective means of overcoming the initial learning curves outside of direct patient interaction initially in the safety and comfort of a laboratory. Such training has been shown to improve technical performance (Grantcharov et al., 2004a), decrease the rate of technical errors (Ahlberg et al., 2007a) and shorten learning curves in the operating room (R. Aggarwal, Ward, et al., 2007). Thus, the major driving force behind the call to incorporate simulation into surgery
training comes from the need to improve patient safety and to shorten the learning curves in the operating room.
2.6 Conclusion

This study used a cohort of international experts and both the classical and modified Delphi methods to develop a consensus on an evidence-based framework for the design, validation, evaluation and implementation of a simulation-based training curriculum for any surgical procedure in any surgical specialty. A step-by-step progression through the proposed framework will provide surgical educators, program directors and decision makers with an evidence-based guide for design and incorporation of a simulation-based surgical training curriculum into their respective residency training programs. Wide implementation of this framework will require the support of national organizations, professional societies and residency review committees.

This curricular framework is the first step to establishing a uniform, interoperable international approach to the full life cycle of curriculum design, development and validation. Further research studies should focus on defining the means for establishing benchmark values for assessment, for defining the appropriate intervals for re-testing of both cognitive and technical skills, and for maintenance practice. Once this is achieved, it will be possible to begin a transition from the current, time-based surgery training paradigm to a proficiency-based paradigm with objective means of certification and recertification. This approach has the potential to improve the overall quality of patient care by standardizing the competencies of surgeons in training and in practice.

Finally, this manuscript and its proposed standardized framework are intended to bring uniformity to the process of curriculum development. This framework is a living document intended for improvement. There are numerous aspects that could have been addressed and added, and will likely be incorporated in the future. However, by providing a “straw man” approach based both on scientific rigor and pragmatic values of implementation, the goal was to achieve a baseline for those intending to develop curricula that would result in interoperability of their efforts with those of the community at large. While the architectural backbone of the framework should be adhered to when possible, the final details do have enough flexibility to be adaptable to many skills, procedures and specialty needs. Careful attention was paid to insuring that this curriculum template can be implemented in most any type of surgery (open, laparoscopic, robotic, endoscopic, image-guided, etc.) as well as across surgical and procedural specialties in order to be able to conduct comparative effectiveness analyses between training and procedures with equivalent parameters, thereby comparing “apples to apples” rather than
“apples to oranges”. As in all guidelines, it is implicit that this framework will change over time and adapt to new discoveries in science, technology, educational principals, etc.
Chapter 3

Development, Feasibility, Validity, and Reliability of a Scale for Objective Assessment of Operative Performance in Laparoscopic Gastric Bypass Surgery

This chapter is modified from the following:

Development, Feasibility, Validity, and Reliability of a Scale for Objective Assessment of Operative Performance in Laparoscopic Gastric Bypass Surgery

3.1 Abstract

**Background**: There is no objective scale for assessment of operative skill in laparoscopic Roux-en-Y gastric bypass (LRYGB). The objective of this study was to develop and demonstrate feasibility of use, validity and reliability of a Bariatric Objective Structured Assessment of Technical Skill (BOSATS) scale.

**Study Design**: The BOSATS scale was developed using a hierarchical task analysis (HTA), a Delphi questionnaire and a panel of international experts in bariatric surgery. The feasibility of use, reliability, and validity of the developed scale was demonstrated by reviewing 52 prospectively collected video recordings of LRYGB performed by novice and experienced surgeons.

**Results**: A total of 214 discrete steps were identified in HTA. A total of 12 and 17 panel members completed the 1st and 2nd round of the Delphi questionnaire. Consensus among the panel was achieved after a 2nd round (Cronbach’s alpha = 0.85). The BOSATS scale demonstrated high inter-rater (ICC = 0.954; p < 0.001) and test-retest reliability (ICC = 0.99; p < 0.001). Significant differences between BOSATS scores of experienced and novice surgeons were noted for the creation of jejunojejunostomy, gastric pouch, linear stapled gastrojejunostomy, circular stapled gastrojejunostomy, and hand sewn gastrojejunostomy. Moderate to high correlations between BOSATS scale and OSATS GRS were seen for creation of the jejunojejunostomy (rho=0.59; p=0.001), gastric pouch (rho=0.48; p=0.0004), linear stapled gastrojejunostomy (rho=0.70; p=0.0001) and hand sewn gastrojejunostomy (rho=0.96; p<0.0001).

**Conclusions**: The BOSATS scale is a feasible to use, reliable and valid instrument for objective assessment of operative performance in LRYGB. Implementation of this scale is expected to facilitate deliberate practice and provide a means for future certification in bariatric surgery.
3.2 Introduction
Over the past two decades, laparoscopic Roux-en-Y gastric bypass (LRYGB) has gained increasing popularity in North America as the procedure of choice for morbid obesity. Over 50,000 LGBP operations were performed in the United States in 2006 (Livingston, 2010). Achieving proficiency in this advanced laparoscopic procedure remains a challenge for both surgery residents and practicing surgeons. With a learning curve of 50-100 cases (Zevin, Aggarwal, & Grantcharov, 2012), surgeons’ technical skill has been shown to correlate strongly with complication rates in bariatric surgery (Birkmeyer, 2012).

Declaration of proficiency in LRYGB is currently based on the number of cases performed as a primary surgeon and/or on a subjective opinion of a senior preceptor (ASMBS, 2012); however, both of these methods are unreliable and imprecise (R. Aggarwal et al., 2008; Reznick, 1993). In an effort to move toward objective measures of surgical proficiency, several evaluation scales have been designed and validated for use in open and laparoscopic surgery. Examples of such scales include the Objective Structured Assessment of Technical Skills Global Rating Scale (OSATS GRS) (J. A. Martin et al., 1997), modified OSATS GRS (Grantcharov et al., 2004a), Global Operative Assessment of Laparoscopic Skills (GOALS) (Vassiliou et al., 2005), as well as procedure-specific assessment scales for laparoscopic Nissen fundoplication (Peyre et al., 2009), and laparoscopic right and sigmoid colectomies (Palter, MacRae, et al., 2011). None of these scales have been developed or validated for use in LGBP.

OSATS GRS, as developed by Reznick and colleagues (J. A. Martin et al., 1997) is probably the most widely utilized form of objective assessment of technical skill in surgery. It was developed and validated for use with a specific set of surgical tasks (bowel anastomosis, vascular anastomosis, etc.) performed on bench top models (J. A. Martin et al., 1997). The GOALS and modified OSATS GRS scales were developed as alternatives to OSATS GRS for laparoscopic surgery; however, both scales were validated only for use in laparoscopic cholecystectomy. OSATS GRS, modified OSATS GRS and GOALS are examples of global rating scales. Such scales, unlike procedure-specific scales, are designed to provide a trainee with a global assessment of their operative performance (respect for tissue, time and motion, etc.). Unfortunately, global rating scales do not provide a trainee with any information on specific parts of an operation that need the most improvement (R. Aggarwal et al., 2008).
At present, there is no objective scale to assess either overall or specific step operative performance in LRYGB. Aggarwal and colleagues were the only group that attempted to address this deficiency by designing and validating a global and procedure-specific assessment scale for the operative step of laparoscopic jejunojunostomy (R. Aggarwal, Boza, et al., 2007). Albeit an excellent start, that scale was designed to assess only one component of the LRYGB operation. Furthermore, it is of limited benefit in live surgery, because it was validated only for use on cadaveric porcine models.

An ideal scale for assessment of surgical proficiency in LRYGB should be comprehensive, feasible to use, internationally relevant, reliable and valid for live and recorded surgery. Such scale should include all common surgical approaches to LRYGB, thereby enabling its use in different institutions and enabling comparability between surgeons. Utilization of such a scale would not only allow for an objective evaluation of the technical proficiency of a specific individual in LRYGB, but also provide a basis for constructive feedback and deliberate practice.

The present study had three objectives. The first objective was to deconstruct the LRYGB procedure into its component steps. The second objective was to use the Delphi method and an international panel of experts in bariatric surgery to select the steps of LRYGB that were considered important for inclusion into a Bariatric Objective Structured Assessment of Technical Skill (BOSATS) scale. The third objective was to demonstrate feasibility of use, inter-rater and test-retest reliability, and construct and concurrent validity for the developed scale.
3.3 Methods

3.3.1 Hierarchical Task Analysis

A hierarchical task analysis (HTA) of LRYGB was carried out to deconstruct the operation into its component steps. A similar approach has been used in the past to define steps for some basic and intermediate laparoscopic procedures (Peyre et al., 2009; Sarker, Chang, Albrani, & Vincent, 2008b). Ten full-length video recordings of LRYGB performed using various surgical approaches (linear stapled jejunojejunostomy [JJ], linear stapled gastrojejunostomy [GJ], circular stapled GJ, hand sewn GJ, antecolic antegastric and retrocolic retrogastric Roux limb placement) were reviewed independently by two experienced laparoscopic surgeons. Each reviewer listed all consecutive steps required for the completion of each operation. An in-person meeting was organized between the reviewers to resolve any disparities in regards to the generated steps. The list of steps for triple-stapled technique for JJ anastomosis and retrocolic, antegastric approach to Roux limb placement was generated by reviewing four videos from two online video databases (www.sages.org; www.websurg.com). As a result, HTA was completed for all possible operative approaches to the LRYGB.
3.3.2 Selection of Steps for Inclusion into the BOSATS Scale

3.3.2.1 Study design

The Delphi method and an online questionnaire were used to achieve consensus among a panel of international experts in bariatric surgery on the steps of LRYGB that were considered of greatest importance for inclusion into the Bariatric Objective Structured Assessment of Technical Skill (BOSATS) scale. The objective for this study was to select steps from the HTA that were of importance for assessment of surgical proficiency in LRYGB - performed using various surgical approaches – and not to achieve expert consensus on the best surgical approach. The Delphi method was used as a means for obtaining opinions from a panel of experts in a systematic manner (Fink et al., 1984; Graham et al., 2003; Nathens et al., 2003; P. L. Williams & Webb, 1994). The responses of the panel were collated, analyzed and presented back in an iterative fashion until consensus was achieved. Research ethics boards of St. Michael’s Hospital and University of Toronto, Toronto, Canada approved the study.

3.3.2.2 Selection of the expert panel

Members of the expert panel were selected based on their roles as international opinion leaders in the field of advanced laparoscopic surgery as evidenced by their roles within North American, South American, European, and Australasian surgical societies. All members of the expert panel were also members of the Advanced Training in Laparoscopic Abdominal Surgery (ATLAS) group. Participation in the study was voluntary, and informed consent was implied if an individual agreed to participate.

3.3.2.3 Administration of the questionnaire

The list of steps for LGBP generated from HTA was presented to each panel member via an online questionnaire (www.surveymonkey.com). Each panel member was asked to rate the operative steps on a Likert-type scale from 1 (“strongly disagree”) to 5 (“strongly agree”) with respect to the degree of relevance for inclusion into the BOSATS scale. Panel members were given an opportunity to comment on their selections and to clarify their ratings. Two questionnaire rounds were administered, and a total of three e-mail reminders for questionnaire completion were sent during each round.
3.3.2.4 Determination of consensus

Cronbach’s alpha was chosen as the statistical index to quantify a measure of consensus among panel members (Graham et al., 2003). Bland and others suggested that Cronbach’s alpha should be above 0.90 for a diagnostic scale to be useful in clinical practice (Bland & Altman, 1997; Shrout & Fleiss, 1979). For educational interventions, alpha is often set at 0.70 (Nunnally, 1978). For this study, alpha $\geq 0.80$ was chosen to represent an acceptable measure of consensus. If consensus was not achieved, group means and standard deviations for each step were calculated, steps were added or modified based on the comments of the panel members, and the questionnaire was resubmitted for rating. This process continued in an iterative fashion until consensus was achieved.

3.3.2.5 Selection of steps for inclusion into the BOSATS scale

Once consensus among the panel members was achieved (alpha $\geq 0.80$), steps that were rated as 4 (“agree”) or 5 (“strongly agree”) by over 80% of panel members were selected for inclusion into the BOSATS scale. Anchoring descriptors for technical proficiency for each step were assigned to a numerical scoring scale ranging from 1 to 5.

3.3.2.6 Statistical Analysis

Descriptive statistics were calculated for each step. Cronbach’s alpha was calculated as a measure of consensus among panel members. Missing data points, which resulted from submissions of incomplete questionnaires, were dealt with in three ways: (1) the missing data point was replaced with a mean for that step; (2) the missing data point was replaced with a 3 (“neutral”) based on the assumption that because the step was omitted, the panel member did not feel strongly about including or excluding that step; and (3) the missing data point was replaced with the mode for that step. Cronbach’s alpha was calculated using each of the above methods.
3.3.3 Determination of Feasibility of Use, Validity and Reliability of the BOSATS Scale

3.3.3.1 Study design

This part of the study used a prospective single blinded observational study design. Video recordings of LRYGB performed by senior surgery residents (postgraduate years 3 – 5), minimally invasive surgery fellows and experienced laparoscopic bariatric surgeons at three academic teaching hospitals in Ontario, Canada were prospectively collected. Only the laparoscopic intra-abdominal camera view was recorded. The following operative approaches to LRYGB were recorded: creation of JJ (linear stapled anastomosis with hand sewn closure of common enterotomy), positioning of the Roux limb (antecolic, antegastric approach), and creation of GJ (circular stapled anastomosis with transoral anvil placement, linear stapled and hand sewn anastomoses). The operative step of obliteration of the mesenteric defects was not included, because the practice patterns of supervising surgeons for this study are not to obliterate the mesenteric defects.

3.3.3.2 Participant selection

Study participants were selected based on case-volume criteria. Novice surgeons were defined as those individuals who had acted as a primary surgeon in fewer than 10 LRYGB operations in its entirety. Minimally invasive surgical fellows who had performed fewer than 10 LRYGB operations were considered to be novices for the purpose of this study. Experienced surgeons were defined as individuals who had performed more than 100 LRYGB procedures as a primary surgeon. Experienced surgeons were expected to be beyond their initial learning curve (Zevin, Aggarwal, et al., 2012). All participants were required to sign an informed consent prior to participation in this study.

3.3.3.3 Video assessments

Two trained and independent raters (B.Z. and E.B.), blinded to the identity of the operating surgeon, reviewed LRYGB video recordings and scored them using the BOSATS scale. If a step on the BOSATS scale was not performed or visualized on the video, the raters were instructed to rate it as “not applicable”. In addition, each discrete component of the operation (creation of JJ, creation of gastric pouch and creation of GJ) was also evaluated using the OSATS GRS scale. This approach was chosen to account for the common practice of novice surgeons learning LRYGB in steps (completion of JJ, completion of GJ, etc.) prior to completing the operation in
its entirety. Furthermore, evaluating the entire procedure with the OSATS GRS would have been methodically incorrect, because procedures were often shared between experienced and novice surgeons.

3.3.3.4 Sample size

A power calculation was performed a priori based on prior work with the global operative assessment of laparoscopic skills (GOALS) (Vassiliou et al., 2005) scale. The minimum relevant difference between novice and experienced groups in that study was 6.4 points. Using a standard deviation of 4.5, power of 0.8 and alpha of 0.05, the minimum required number of videos for each group was 8.

3.3.3.5 Feasibility of use

All minimally invasive operating rooms at the academic hospitals involved in this study were equipped with video-recording equipment. It is the standard of practice at these hospitals to record all laparoscopic cases. As a result, there were no associated costs with obtaining the video-recordings. In regards to the utilization of the BOSATS scale, raters were asked to record the average duration of time required to score a LRYGB video recording. Fast-forwarding of video recordings was permitted.

3.3.3.6 Reliability

Inter-rater reliability for the BOSATS scale was calculated by correlating total and component scores between two independent raters using the Intraclass Correlation Coefficient, ICC (two-way mixed-effects model, absolute agreement). The use of two independent raters and Intraclass Correlation Coefficient for calculation of reliability of an assessment scale in surgery has been deemed acceptable by several authors (Van Nortwick et al., 2010; Weir, 2005). Test-retest reliability for the BOSATS scale was calculated by correlating the total score for case # 1 assessed in its entirety at two different time points using ICC. The reliability of the OSATS GRS has been demonstrated in a prior study (J. A. Martin et al., 1997), therefore it was not necessary to replicate these calculations in the present study.

3.3.3.7 Validity

Differences between novice and experienced surgeon group scores were analysed to provide evidence of construct validity for each of the following components: creation of JJ, creation of
gastric pouch and creation of GJ. Surgeons were assigned to a novice or an experienced group based on previously defined case-volume criteria. In view of the small sample size, data were treated as non-parametric. Scores of two raters for each component were pooled together and the complete data set was analysed using a Mann-Whitney U test to assess differences between groups. Bonferroni correction was applied to account for a potential for repeated measures in view of pooling rater scores. Statistical significance was set to P < 0.025. Group scores were reported as median (interquartile range).

Data was then reanalyzed using the following definitions for novice and experienced surgeons: surgeons that scored < 28 on the OSATS GRS scale were assigned to the novice group, and surgeons that scored >/= 28 were assigned to the experienced group. This method of grouping avoided the inherent subjectivity associated with classification of surgeons based on their case-volumes.

Concurrent validity for creation of JJ, gastric pouch and creation of GJ components was calculated by correlating BOSATS scores with OSATS GRS scores using Spearman’s correlation coefficient.

3.3.3.8 Statistical analysis

All statistical analyses were performed using STATA version 12.0 (College Station, Texas, USA).
3.4 Results

3.4.1 Hierarchical Task Analysis

A total of 14, full-length video recordings of LRYGB performed using the following common approaches were reviewed: creation of JJ anastomosis (linear stapled anastomosis with hand sewn or stapled closure of common enterotomy), placement of the Roux limb (antecolic antegastric, retrocolic retrogastric, retrocolic antegastric), creation of GJ anastomosis (circular stapled with transgastric and transoral anvil placement, linear stapled, hand sewn). A total of 214 discrete steps were identified for all examined approaches (Table 8).

3.4.2 Selection of Steps for Inclusion into the BOSATS Scale

Twelve international experts in laparoscopic bariatric surgery from five countries were invited to complete the 1st round of the online Delphi questionnaire. All 12 individuals (100%) completed the questionnaire (Table 9). Cronbach’s alpha for the entire questionnaire (containing all steps from HTA) was 0.64 when missing data points were replaced by mean, 0.55 when missing data points were replaced by 3 “neutral”, and 0.62 when missing data points were replaced by mode. Cronbach’s alpha for individual components of the questionnaire ranged from 0.25 to 1.00 (Table 10).

A total of 19 international experts in laparoscopic bariatric surgery from seven countries were invited to complete the 2nd round of the online Delphi questionnaire. Twelve of the 19 experts were the same individuals utilized for the 1st round of the Delphi questionnaire. Seven of the 19 experts were added to the group for 2nd round. Seventeen participants (89%) from six countries completed the survey (Table 9). Cronbach’s alpha for the entire questionnaire was 0.85, 0.83 and 0.84 when missing data points were replaced by mean, 3 “neutral” and mode, respectively. Cronbach’s alpha for individual components of questionnaire ranged from 0.15 to 0.97 (Table 10). Once consensus among panel members was achieved, a total of 99 out of 214 (46%) steps from the questionnaire were selected for inclusion into the BOSATS scale (Table 11).
Table 8. List of all possible operative steps for the LRYGB.

INSERT UMBILICAL PORT
- Decide on location of port
- Local anaesthesia
- Transverse 12mm incision to left of umbilicus
- Set up XL port
- Focus camera
- Insert port to peritoneal cavity
- Push port in a little more
- Remove camera and introducer
- Attach gas - 14mm Hg

DIAGNOSTIC LAPAROSCOPY
- Scope to L side
- Scope to pelvis
- Back to epigastrium
- Scope to R side

PLACE NATHANSON'S RETRACTOR
- Local anaesthesia
- Make 5mm transverse incision in epigastrium
- Insert 5mm introducer
- Insert retractor through abdominal wall
- Place retractor under liver
- Fix retractor to holder

PLACE THREE FURTHER PORTS
- Local anaesthesia
- Make 12mm incision at left flank (anterior axillary line)
- Local anaesthesia
- Make 12mm incision at left side (mid clavicular line)
- Local anaesthesia
- Make 12mm incision at right of midline
- Insert left flank port
- Insert left side port
- Insert right of midline port
- Move gas to left side port
- Attach smoke retractor to left flank port
- Give camera to assistant

DISSECT ANGLE OF HIS
- Place patient in steep reverse transdelenberg
- Harmonic L hand
- Babcock R hand
- Pull fundus of stomach down
- Might need Babcock assistant to push away omentum
- Dissect angle close to stomach
- Keep tension with R hand

GASTRIC POUCH
- Empty stomach by pressing on fundus with R hand Babcock
  (Anaesthetist to suction NG tube)
- Assistant Babcock on lesser omentum - push down and to left of screen
- R hand Babcock pull lesser curve of stomach up
- Dissect peri-gastric area with L hand Harmonic
- Keep close to stomach
- Make posterior tunnel - insert R hand Babcock
- Stapler blue L hand
- Push omentum away with R hand Babcock
Fire stapler
Attend to bleeding - Harmonic R hand
Babcock L hand and Harmonic R hand
Pull stomach up with L hand
Mobilise posterior attachments to gastric pouch
Continue tunnel to Angle of His
Assistant Babcock in tunnel
Assistant Babcock to come through window in Angle of His
L hand Babcock to retract pouch down and to L of screen
Stapler blue R hand
Staple fire
Retract pouch down and to L of screen - Babcock L hand
Staple fire

DIVIDE GREATER OMENTUM
Johan L hand
Harmonic R hand
Assistant Babcock to hold omentum to right
Harmonic to divide omentum to transverse colon
Extend either side at T

IDENTIFY AND MOBILISE JEJUNUM FOR BP LIMB
Take patient out of steep reverse trandelengerg
Assistant Babcock under transverse colon
Johan L hand
Babcock right hand
Identify ligament of Treitz
Measure 50cm small bowel distal to the ligament
BP loop to R side of screen
Assistant holds small bowel with Babcock

GASTRO-JEJUNAL ANASTOMOSIS (LINER STAPLER)
Johan L hand
Harmonic R hand - gastrotomy in centre of staple line
Johan L hand
Hold SB to left
Harmonic – enterotomy on BP limb
Stapler (narrow part) into jejunum
Assistant lets go
Stapler (wide part) into pouch
Staple
Retract stapler
Suture on R side - stay stitch - backhand
Insert a 34 F tube across anastomosis to ensure patency while suturing
Johan L hand
Needle holder right hand
Suture on L side - to R side stitch - forehand
Assistant to hold suture in between stitches
Tie to R side short end
Harmonic R hand to cut sutures
Remove needles (x2) from abdomen

Alternatively can staple the enterotomy with a linear stapler

GASTRO-JEJUNAL ANASTOMOSIS (CIRCULAR STAPLER)
Pass nasogastric tube into the stomach with anvil attached
Make a gastrotomy in the gastric pouch
Pull nasogastric tube and attached anvil into gastric pouch
Use Harmonic to make an enterotomy at the staple line of Roux limb to introduce the circular stapler
Dilate LUQ port site
Insert 21 mm EEA circular stapler into Roux limb and advance 7 cm
Perforate the wall of Roux limb with spike and connect stapler to anvil
Fire stapler
Remove stapler
Replace port
Close Roux limb enterotomy with 2.5mm liner stapler or single layer running suture

ALTERNATIVELY
Insert a wire through an angiocatheter placed through the abdominal wall in LUQ
Insert an endoscope into the gastric pouch with a snare device
Make a small gastrotomy with Harmonic
Retrieve the wire with the snare and pull endoscope out through the mouth
Attach the anvil to the wire and pull it back into the gastric pouch
Cut the wire off the anvil
Remove wire (x1) from abdomen
Use Harmonic to make enterotomy distal to staple line of Roux limb
Dilate LUQ port site
Insert 21 mm EEA circular stapler via port site into Roux limb
Connect stapler to anvil
Fire stapler
Remove stapler
Replace port
Close enterotomy with 2.5mm liner stapler or single layer running suture

GASTRO-JEJUNAL ANASTOMOSIS (HAND SEWN)
Create a posterior seromuscular running suture line (sew towards yourself) between Gastric Pouch and Roux limb
Cut end of suture long
Remove needle (x1) from abdomen
Create and extend 10-15mm enterotomies on the gastric pouch and the Roux limb
Place a second posterior layer sewing towards the camera full thickness running suture
Insert a 34 F gastric tube and feed it across the anastomosis
Create an anterior full thickness running suture line over the gastric tube using the same stitch
Tie to the origin and cut
Remove needle (x1) from abdomen
Place a second anterior running seromuscular suture (sewing towards the camera) and tie to the first suture that was left long
Cut and remove suture (x1) from abdomen

ROUX LIMB POSITIONING (ANTECOLIC, ANTIGASTRIC)
DIVIDE GREATER OMENTUM
Johan L hand
Harmonic R hand
Assistant Babcock to hold omentum to right
Harmonic to divide omentum to transverse colon
Extend either side at T
Bring up the Roux Limb antecolic and antigastric
Prepare to begin the gastro-jejunostomy

ROUX LIMB POSITIONING (ANTECOLIC, RETROGASTRIC)
DIVIDE GREATER OMENTUM
Johan L hand
Harmonic R hand
Assistant Babcock to hold omentum to right
Harmonic to divide omentum to transverse colon
Extend either side at T

ROUX LIMB POSITIONING (RETROCOLIC, ANTIGASTRIC)
Grasp Transverse Mesocolon with 2 atraumatic graspers
Make an incision in the mesocolon with Harmonic scalpel to the left of the middle colic vessels
Carry incision cephalad until posterior aspect of stomach is identified
Pull Roux limb through the mesocolic defect and position in the antigastric position

**ROUX LIMB POSITIONING (RETROCOLIC, RETROGASTRIC)**
Grasp Transverse Mesocolon with 2 atraumatic graspers
Make an incision in the mesocolon with Harmonic scalpel to the left of the middle colic vessels
Carry incision cephalad until posterior aspect of stomach is identified
Pull Roux limb through the mesocolic defect and position in the retrogastric position

**MOBILISE JEJUNUM FOR ALIMENTARY LIMB**
Johan L hand
Babcock R hand
Bring loop up - remainder to left side of screen
Push remainder to L of screen and down
Bowel to assistant - Babcock

**JEJUNO-JEJUNOSTOMY**
L hand Johan to hold bowel above
Enterotomy R hand harmonic scalpel
L hand Johan to bowel
Insert wide limb of stapler R hand
Assistant let go
Insert staple limb to other side
Pull with Johan
Staple fire
Staple fire other way - L hand
Johan R hand
Stay suture to bottom of enterotomy
Johan L hand
Needle holder R hand
Suture from top to bottom
Johan L hand
Needle holder R hand
Assistant to hold suture in between stitches
Harmonic R hand to cut sutures
Remove both needles from abdomen

**CLOSE SB MESENTERIC DEFECT**
Needle holder right hand - Ethibond
Johan L hand
First stitch
Tie stitch
Running stitch down and then back up
Lock stitches
Assistant to hold suture tight in between stitches
Tie to original short end
Cut suture with Harmonic - R hand
Remove needle L hand

**DIVIDE OMEGA LOOP**
Harmonic L hand
Johan R hand
Lift bowel with R hand
Divide omentum
Johan through window
Stapler L hand
Push through
Fire stapler
Pull BP limb down

**CLOSE PETERSEN'S DEFECT**
Johan L hand
Needle holder right hand
Assistant Babcock to pull transverse colon up
Stitch between SB mesentery and mesocolon
Tie knot
Stitch between SB mesentery and fat pad
Tie knot
Harmonic R hand
Cut suture
Remove needle L hand
Babcock L hand
Johan R hand
Place SB to L of screen

GASTRO-RJEJUNAL LEAK TEST
Babcock L hand
Clamp just below GJ anastomosis
Pull down
Suction R hand
Fill gastric pouch with 120 cc of methylene blue dye
Suction instrument to L and R corners
Suck back fluid from gastric pouch

ALTERNATIVELY
Introduce an endoscope into gastric pouch
Instil water into peritoneal cavity
Submerge the gastro-jejuno-stomy
Occlude Roux-limb distal to GJ anastomosis with clamp
Insufflate the gastric pouch with air
Check for air leak

REMOVE NATHANSON'S RETRACTOR
Unscrew
Remove in one movement

RETRACT PORTS UNDER VISION
L flank
L side
Right side

REMOVE INTRA-ABDOMINAL GAS
Remove cover for umbilical port
Pull up port
Open valves for other three 12mm ports

CLOSE WOUNDS
Two sharp towel clips - one on either end
Clip applier to assistant
3 to 12mm ports
1 to 5mm port
Table 9. Composition of the international expert panel for the online Delphi questionnaire.

<table>
<thead>
<tr>
<th>Location</th>
<th>Participants in first round, n</th>
<th>Participants in the second round, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contacted</td>
<td>Responded</td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>USA</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>UK</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chile</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Australia</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sweden</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Response rate:</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
Table 10. Agreement among the Delphi panel members for the 1st and 2nd round of online Delphi questionnaire.

<table>
<thead>
<tr>
<th>Component of the questionnaire</th>
<th>First round</th>
<th>Second round</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>alpha</td>
</tr>
<tr>
<td>Entire questionnaire (all components included)</td>
<td>12</td>
<td>0.55</td>
</tr>
<tr>
<td>Patient positioning</td>
<td>12</td>
<td>0.89</td>
</tr>
<tr>
<td>Abdominal access</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Veress needle technique</td>
<td>11</td>
<td>0.68</td>
</tr>
<tr>
<td>Optical trocar without prior insufflation</td>
<td>12</td>
<td>0.62</td>
</tr>
<tr>
<td>Diagnostic laparoscopy and port placement</td>
<td>12</td>
<td>0.74</td>
</tr>
<tr>
<td>Placement of liver retractor</td>
<td>12</td>
<td>0.90</td>
</tr>
<tr>
<td>Creation of the Roux limb</td>
<td>12</td>
<td>0.56</td>
</tr>
<tr>
<td>Jejunojejunostomy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear stapled with hand-sewn or stapled closure of common enterotomy</td>
<td>12</td>
<td>0.81</td>
</tr>
<tr>
<td>Triple stapling technique</td>
<td>11</td>
<td>0.61</td>
</tr>
<tr>
<td>Dissection of the phreno-esophageal ligament</td>
<td>12</td>
<td>0.89</td>
</tr>
<tr>
<td>Creation of gastric pouch</td>
<td>12</td>
<td>0.64</td>
</tr>
<tr>
<td>Roux limb positioning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antecolic, antegastric</td>
<td>12</td>
<td>0.75</td>
</tr>
<tr>
<td>Retrocolic, retrogastric</td>
<td>9</td>
<td>1.00</td>
</tr>
<tr>
<td>Retrocolic, antegastric</td>
<td>10</td>
<td>0.25</td>
</tr>
<tr>
<td>Creation of gastrojejunostomy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear stapled anastomosis with hand-sewn closure of common enterotomy</td>
<td>9</td>
<td>0.81</td>
</tr>
<tr>
<td>Circular stapled anastomosis</td>
<td>11</td>
<td>0.63</td>
</tr>
<tr>
<td>Hand-sewn anastomosis</td>
<td>11</td>
<td>0.58</td>
</tr>
<tr>
<td>Closure of mesenteric defects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jejunojejunostomy defect</td>
<td>12</td>
<td>0.85</td>
</tr>
<tr>
<td>Petersen’s defect</td>
<td>11</td>
<td>0.49</td>
</tr>
<tr>
<td>Transverse mesocolon defect</td>
<td>9</td>
<td>0.74</td>
</tr>
<tr>
<td>Gastrojejunostomy leak test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blue dye</td>
<td>12</td>
<td>0.36</td>
</tr>
<tr>
<td>Air insufflation</td>
<td>10</td>
<td>1.00</td>
</tr>
<tr>
<td>Removal of ports and retractors</td>
<td>12</td>
<td>0.90</td>
</tr>
<tr>
<td>Wound closure</td>
<td>12</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Missing data points were replaced by 3 “neutral” for analysis. Data reported as n, number of participants responding; alpha denotes Cronbach’s alpha.
<table>
<thead>
<tr>
<th>Task / Step</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient Positioning:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supine with pressure points padded</td>
<td>Not performed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perform promptly, skillfully, without prompting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Access and Port Insertion:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Veress needle</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insert Veress needle in the left subcostal area midclavicular line</td>
<td>Inserted at an incorrect angle; incorrect location; failed to enter peritoneal cavity</td>
<td>Inserted at a correct angle and correct anatomic location; entered peritoneal cavity; some hesitation</td>
<td>Inserted at a correct angle, correct location, smooth and skillful</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish pneumoperitoneum (15-18 mmHg)</td>
<td>Not performed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perform promptly with correct pressure setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Insert a camera port using an optical viewing trocar</strong></td>
<td>Performed clumsily and with difficulty; inadequate size of skill incision</td>
<td>Performed adequately; partial peritoneal penetration</td>
<td>Performed smoothly and skillfully</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confirm absence of injuries from Veress needle</td>
<td>Not performed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate confirmation of Veress position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Optical viewing trocar without prior abdominal insufflation</strong></td>
<td>Not clear on the appropriate location for the trocar</td>
<td>Appropriate location for the trocar selected after a delay and with prompting</td>
<td>Appropriate location selected without prompting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decide on location of the optical viewing trocar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Make a skin incision and introduce optical viewing trocar into peritoneum</strong></td>
<td>Incision made in an incorrect location; extensive tissue trauma; too small; trocar inserted with difficulty</td>
<td>Incision of adequate size; made tentatively, although in correct location; some tissue trauma; adequate trocar insertion</td>
<td>Incision of adequate size; correct location; smooth insertion without excessive tissue trauma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confirm trocar placement by visualizing the layers of the abdominal wall</td>
<td>Not done</td>
<td>Confirmed with some difficulty</td>
<td>Clearly confirmed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Remove obturator and verify that the trocar is in the abdominal cavity prior to insufflation</strong></td>
<td>Not removed</td>
<td>Removed after a delay</td>
<td>Removed promptly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish pneumoperitoneum (15 – 18 mm Hg)</td>
<td>Not performed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perform promptly and with correct pressure setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confirm absence of injuries from trocar insertion</strong></td>
<td>Not performed</td>
<td>Injury could not be definitively excluded</td>
<td>Adequate confirmation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perform diagnostic laparoscopy</td>
<td>Not performed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Careful and thorough inspection of all 4 quadrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Placement of additional ports</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place 3-4 additional ports under direct</td>
<td>Performed clumsily; skiving; poor choice of port location; no direct</td>
<td>Performed adequately; no skiving</td>
<td>Performed smoothly and skillfully</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11. The Bariatric Objective Structured Assessment of Technical Skill (BOSATS) scale.
### Placement of Liver Retractor:

<table>
<thead>
<tr>
<th>Insert liver retractor through the abdominal wall and place under the liver to expose the diaphragmatic hiatus</th>
<th>Inserted in incorrect position; poor orientation; poor exposure</th>
<th>Inserted correctly, but required repositioning; adequate exposure</th>
<th>Inserted in a smooth, controlled motion; excellent exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fix liver retractor to holder</td>
<td>Not performed</td>
<td>Performed after some delay or with prompting</td>
<td>Performed quickly and skillfully</td>
</tr>
</tbody>
</table>

### Creation of the Roux Limb:

<table>
<thead>
<tr>
<th>Elevate the transverse colon cephalad and identify ligament of Treitz</th>
<th>Not performed</th>
<th>Performed in a traumatic fashion; ligament identified</th>
<th>Elevated smoothly and gently; good exposure of ligament</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure approximately 40-60cm of jejunum distal to the ligament of Treitz</td>
<td>Length not measured</td>
<td>Measured, however individual measurements not of the same size; poor orientation</td>
<td>Measured methodologically; each measurement of the same size; correct orientation</td>
</tr>
<tr>
<td>Confirm that this part of jejunum will reach the hiatus / gastric pouch</td>
<td>Not confirmed</td>
<td>Confirmed briefly</td>
<td>Confirmed clearly and methodologically</td>
</tr>
<tr>
<td>Divide jejunum</td>
<td>Divided with tissue trauma and gross contamination</td>
<td>Bowel divided; jaws of stapler not perpendicular to bowel; possibility of mesenteric trauma; need for re-resection</td>
<td>Bowel divided at correct angle; good viability of ends; no trauma to mesentry</td>
</tr>
</tbody>
</table>

### Creation of JJ anastomosis:

#### Preparation

<table>
<thead>
<tr>
<th>Measure 75-150 cm length of Roux limb</th>
<th>Length not measured</th>
<th>Measured, however individual measurements not of the same size; poor orientation; grabs onto mesentery</th>
<th>Measured methodologically; each measurement of the same size; correct orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bring distal end of BP limb and position side to side to Roux limb</td>
<td>Incorrect orientation of BP limb to Roux limb</td>
<td>Correct orientation, although required multiple attempts with extra movements</td>
<td>BP limb positioned correctly in relation to Roux limb; no extra movements; no tension</td>
</tr>
<tr>
<td>Create enterotomies in BP and Roux limbs</td>
<td>Poor relation between grasper and energy source; excessively large or small; penetration of posterior bowel wall</td>
<td>Appropriate size enterotomy; not placed in antimesenteric location</td>
<td>Appropriately sized and placed enterotomies; no extra movements. Good relation of grasper and energy source</td>
</tr>
</tbody>
</table>

#### Stapling

<table>
<thead>
<tr>
<th>Insert the limbs of linear cutting stapler into the enterotomies in Roux and BP limbs</th>
<th>Unclear of how to insert the staple device. Drives staple jaws blindly into BP and Roux limb</th>
<th>Inserts the stapler with hesitation and lacks appreciation of the ideal angle for insertion</th>
<th>Inserts staple jaws with ease; controlled manner; correct angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure both limbs are symmetrical and stapler on anti-mesenteric border</td>
<td>Does not ensure limb symmetry and anti-mesenteric position prior to closure of jaws</td>
<td>Limbs either non-symmetrical or not on anti-mesenteric border prior to closure of jaws</td>
<td>Correctly ensures symmetry and anti-mesenteric position prior to closure of the jaws</td>
</tr>
<tr>
<td>Fire stapler</td>
<td>Uncontrolled fire with excessive pull on the bowel and widening of enterotomies</td>
<td>Controlled fire; some slippage of bowel from jaws</td>
<td>Smooth, controlled fire; no widening of enterotomies</td>
</tr>
<tr>
<td>Close the created enterotomy with a simple running suture or a linear cutting stapler</td>
<td>Poorly positioned stitch or stapler. Blindly placed sutures; traumatic use of needle drivers; skiving and poor control of needle; too much or too little tension; poor quality knot</td>
<td>Adequate stitch or stapler position. Adequate closure of enterotomy. Sutures placed at varying distances apart; gathering of bowel edges; additional reinforcement sutures required; stitch pulled out clumsily.</td>
<td>Correct stitch or stapler position. Adequate size bites placed uniform distance apart; perpendicular to seromuscular layer; appropriate tension; adequate surgical knot; needle retrieved safely</td>
</tr>
</tbody>
</table>

#### Triple-stapling technique
<table>
<thead>
<tr>
<th>Insert the limbs of linear cutting stapler into the enterotomies in Roux and BP limbs</th>
<th>Unclear of how to insert the staple device. Drives staple jaws blindly into BP and Roux limb</th>
<th>Inserts the stapler with hesitation and lacks appreciation of the ideal angle for insertion</th>
<th>Inserts staple jaws with ease; controlled manner; correct angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure both limbs are symmetrical and stapler on anti-mesenteric border</td>
<td>Does not ensure limb symmetry and anti-mesenteric position prior to closure of jaws</td>
<td>Limbs either non-symmetrical or not on anti-mesenteric border prior to closure of jaws</td>
<td>Correctly ensures symmetry and anti-mesenteric position prior to closure of the jaws</td>
</tr>
<tr>
<td>Fire stapler</td>
<td>Uncontrolled fire with excessive pull on the bowel and widening of enterotomies</td>
<td>Controlled fire; some slippage of bowel from jaws</td>
<td>Smooth, controlled fire; no widening of enterotomies</td>
</tr>
<tr>
<td>Place linear reticulating stapler through the common enterotomy at 180 degrees to the previous anastomosis</td>
<td>Unclear of how to insert the staple device. Drives staple jaws blindly into the common enterotomy</td>
<td>Inserts the stapler with hesitation and lacks appreciation of the ideal angle for insertion</td>
<td>Inserts staple jaws with ease; controlled manner; correct angle</td>
</tr>
<tr>
<td>Ensure stapler is on the anti-mesenteric border</td>
<td>Does not anti-mesenteric placement prior to closure of jaws</td>
<td>Stapler placed on the anti-mesenteric border after some repositioning</td>
<td>Correctly placed stapler in anti-mesenteric position</td>
</tr>
<tr>
<td>Fire stapler</td>
<td>Uncontrolled fire with excessive pull on the bowel and widening of enterotomies</td>
<td>Controlled fire; some slippage of bowel from jaws</td>
<td>Smooth, controlled fire; no widening of enterotomies</td>
</tr>
<tr>
<td>Apply a liner cutting stapler transversely to the enterotomy</td>
<td>Stapler applied in an incorrect orientation</td>
<td>Stapler applied transversely after multiple repositioning attempts</td>
<td>Stapler applied in correct orientation</td>
</tr>
<tr>
<td>Fire stapler to close the enterotomy</td>
<td>Uncontrolled fire with excessive pull on the bowel</td>
<td>Controlled fire; some slippage of bowel</td>
<td>Smooth, controlled fire</td>
</tr>
</tbody>
</table>

**Dissection of the Phreno-Esophageal Ligament (Angle of HIS):**

<table>
<thead>
<tr>
<th>Place patient in steep reverse Trendelenburg</th>
<th>Patient not repositioned</th>
<th>Patient repositioned after prompting</th>
<th>Patient repositioned without delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pull fundus of stomach down (exposure)</td>
<td>Insufficient retraction; traumatic; insufficient exposure</td>
<td>Satisfactory retraction after some repositioning; sub-optimal exposure</td>
<td>Appropriate retraction; optimal exposure</td>
</tr>
<tr>
<td>Create a tunnel between the left crus of diaphragm and fundus of the stomach</td>
<td>Incorrect location; associated trauma, bleeding</td>
<td>Correct location after multiple attempts; minimal tissue trauma, bleeding</td>
<td>Correct location; no difficulty or excessive tissue trauma</td>
</tr>
<tr>
<td>Dissect angle of HIS close to stomach while keeping tension on fundus</td>
<td>Dissection in incorrect plane; insufficient or too much tension; bleeding</td>
<td>Dissection in correct plane; appropriate tension majority of the time; occasional tissue damage, bleeding</td>
<td>Dissection in correct plane; careful handling of tissues; appropriate tension at all times; minimal tissue damage, bleeding</td>
</tr>
</tbody>
</table>

**Creation of the Gastric Pouch:**

<table>
<thead>
<tr>
<th>Dissect along lesser curvature of stomach 3-8 cm from the GE junction and keep close to stomach</th>
<th>Incorrect plane; incorrect anatomic location; excessive tissue trauma; bleeding with need for suction</th>
<th>Correct plane developed with some difficulty; moderate tissue trauma; bleeding not requiring suction</th>
<th>Correct plane in correct anatomic location developed without difficulty or excessive tissue trauma, bleeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create a posterior tunnel</td>
<td>Dissection in incorrect plane; unnecessary force; bleeding requiring suction</td>
<td>Dissection in correct plane; occasional tissue damage; bleeding not requiring suction</td>
<td>Dissection in correct plane; careful handling of tissues, minimal tissue damage, bleeding</td>
</tr>
<tr>
<td>Introduce and apply a linear cutting stapler transversely to the stomach</td>
<td>Stapler applied in incorrect orientation; serosal damage to stomach</td>
<td>Stapler applied transversely after multiple repositioning attempts</td>
<td>Stapler applied transversely; no requirement for multiple repositioning attempts; no trauma to stomach wall</td>
</tr>
<tr>
<td>Remove all tubes from the stomach prior to firing the stapler</td>
<td>Not done</td>
<td>Done after delay; with prompting</td>
<td>Done without delay</td>
</tr>
<tr>
<td>Fire stapler</td>
<td>Uncontrolled fire with excessive pull on the stomach</td>
<td>Controlled fire; some slippage of stomach between jaws</td>
<td>Smooth, controlled fire</td>
</tr>
<tr>
<td>Step</td>
<td>Correct Technique</td>
<td>Incorrect Technique</td>
<td>Possible Results</td>
</tr>
<tr>
<td>------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Develop a posterior tunnel towards the angle of His</td>
<td>Dissection in incorrect plane; unnecessary force; bleeding requiring suction</td>
<td>Dissection in correct plane; occasional tissue damage; bleeding not requiring suction</td>
<td>Dissection in correct plane; careful handling of tissues, minimal tissue damage, bleeding</td>
</tr>
<tr>
<td>Clean up posterior attachments (if present and prevent introduction of stapler) between the stomach and pancreas in the lesser sac</td>
<td>Posterior attachments not cleared; excessive force, tissue trauma, bleeding requiring suction</td>
<td>Majority of attachments cleared up; moderate tissue trauma, bleeding not requiring suction</td>
<td>Attachments cleared; careful handling of tissues; minimal tissue trauma, bleeding</td>
</tr>
<tr>
<td>Introduce and apply another linear cutting stapler to the stomach</td>
<td>Stapler applied in an incorrect orientation; serosal damage to stomach</td>
<td>Stapler applied correctly; multiple repositioning attempts</td>
<td>Stapler applied incorrectly; no repositioning required; no trauma to stomach wall</td>
</tr>
<tr>
<td>Fire stapler</td>
<td>Uncontrolled fire with excessive pull on the stomach</td>
<td>Controlled fire; some slippage of stomach between jaws</td>
<td>Smooth, controlled fire</td>
</tr>
<tr>
<td>Confirm complete transection of stomach</td>
<td>Not confirmed</td>
<td>Confirmed briefly without adequate visualization</td>
<td>Methodical confirmation of complete transection</td>
</tr>
</tbody>
</table>

### Positioning of the Roux limb:

**Antecolic, Antegastric**

<table>
<thead>
<tr>
<th>Step</th>
<th>Correct Technique</th>
<th>Incorrect Technique</th>
<th>Possible Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>If required, divide omentum to the transverse colon with an energy source</td>
<td>Not performed when necessary; excessive tissue trauma; bleeding; injury to colon</td>
<td>Performed appropriately; moderate tissue trauma; bleeding</td>
<td>Performed correctly; minimal tissue trauma; bleeding</td>
</tr>
<tr>
<td>Bring up Roux limb antecolic and antegastric</td>
<td>Incorrect position; excessive tension or twisting of limb</td>
<td>Correct positioning; repositioning required; limb slips</td>
<td>Correct positioning; minimal tension or twisting</td>
</tr>
</tbody>
</table>

**Retrocolic, Retrogastric**

<table>
<thead>
<tr>
<th>Step</th>
<th>Correct Technique</th>
<th>Incorrect Technique</th>
<th>Possible Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grasp transverse mesocolon with an atraumatic grasper</td>
<td>Multiple attempts to grasp; incorrect instrument used; excessive tissue trauma; bleeding requiring suction</td>
<td>Correct instruments; multiple attempts; moderate tissue trauma; bleeding not requiring suction</td>
<td>Grasped gently; correct instruments; minimal tissue trauma or bleeding</td>
</tr>
<tr>
<td>Incise the mesocolon to the left of the middle colic vessels and carry incision cephalad until posterior aspect of stomach is identified</td>
<td>Incorrect anatomic location; incorrect plane; posterior stomach not identified;</td>
<td>Correct anatomic location and plane; moderate tissue trauma, bleeding</td>
<td>Correct anatomic location and plane; excellent hemostasis</td>
</tr>
<tr>
<td>Pull Roux limb through the mesocolic defect and position in the retrogastric position; avoid twisting of the limb</td>
<td>Incorrect position; excessive tension or twisting of limb</td>
<td>Correct positioning; repositioning required; limb slips</td>
<td>Correct positioning; minimal tension or twisting</td>
</tr>
</tbody>
</table>

**Retrocolic, Antegastric**

<table>
<thead>
<tr>
<th>Step</th>
<th>Correct Technique</th>
<th>Incorrect Technique</th>
<th>Possible Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Divide the gastrocolic ligament</td>
<td>Incorrect plane; excessive tissue trauma; bleeding requiring suction</td>
<td>Performed adequately; correct plane; some tissue trauma or bleeding</td>
<td>Performed correctly; correct plane; minimal tissue trauma or bleeding</td>
</tr>
<tr>
<td>Grasp transverse mesocolon with an atraumatic grasper</td>
<td>Multiple attempts to grasp; incorrect instrument used; excessive tissue trauma; bleeding requiring suction</td>
<td>Correct instruments; multiple attempts; moderate tissue trauma; bleeding not requiring suction</td>
<td>Grasped gently; correct instruments; minimal tissue trauma or bleeding</td>
</tr>
<tr>
<td>Incise the mesocolon to the left of the middle colic vessels and carry incision cephalad until posterior aspect of stomach is identified</td>
<td>Incorrect anatomic location; incorrect plane; posterior stomach not identified;</td>
<td>Correct anatomic location and plane; moderate tissue trauma, bleeding</td>
<td>Correct anatomic location and plane; excellent hemostasis</td>
</tr>
<tr>
<td>Pull Roux limb through the mesocolic defect and window in the gastrocolic ligament; position it in an antegastric position; avoid twisting of the limb</td>
<td>Incorrect position; excessive tension or twisting of limb</td>
<td>Correct positioning; repositioning required; limb slips</td>
<td>Correct positioning; minimal tension or twisting</td>
</tr>
</tbody>
</table>

### Creation of GJ anastomosis:

#### Linear Stapler Technique

<table>
<thead>
<tr>
<th>Step</th>
<th>Correct Technique</th>
<th>Incorrect Technique</th>
<th>Possible Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create a gastrotomy in the gastric pouch</td>
<td>No entry into gastric lumen; poor relation between grasper and energy source; excessively large or small; penetration of posterior bowel wall;</td>
<td>Entry into gastric lumen; appropriate size; more than 1 attempt required</td>
<td>Entry into gastric lumen; appropriate size; no extra movements required</td>
</tr>
<tr>
<td>Action</td>
<td>Observations</td>
<td>Comments</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Create an enterotomy in the Roux limb</td>
<td>No entry into bowel lumen; poor relation between grasper and energy source; excessively large or small; penetration of posterior bowel wall</td>
<td>Appropriate size and entry into bowel lumen; not placed in antimesenteric location</td>
<td>Appropriate size and placement of enterotomy; good relation of grasper and energy source; no extra movements required</td>
</tr>
<tr>
<td>Introduce one limb of linear cutting stapler into gastric pouch and the other into Roux limb</td>
<td>Unclear of how to insert the staple device; drives staple jaws blindly into the enterotomies</td>
<td>Inserts the stapler, but lacks appreciation of the ideal angle for insertion</td>
<td>Inserts staple jaws with ease; controlled manner; correct angle</td>
</tr>
<tr>
<td>Ensure both limbs are symmetrical before firing the stapler</td>
<td>Does not ensure symmetry anti-mesenteric location of stapler prior to closing of jaws</td>
<td>Limbs either non-symmetrical or not on anti-mesenteric border prior to closure of jaws</td>
<td>Correct symmetry and anti-mesenteric position prior to closure of the jaws</td>
</tr>
<tr>
<td>Fire stapler</td>
<td>Uncontrolled fire with excessive pull on the bowel and widening of enterotomies</td>
<td>Controlled fire; some slippage of bowel from jaws</td>
<td>Smooth, controlled fire; no widening of enterotomies</td>
</tr>
<tr>
<td>Suture closed the enterotomy</td>
<td>Poorly positioned stitch; blindly placed sutures; traumatic use of needle drivers; skiving and poor control of needle; too much or too little tension; poor quality knot</td>
<td>Adequate stitch position; adequate closure of enterotomy. Sutures placed at varying distances apart; gathering of bowel edges; additional reinforcement sutures required; stitch pulled out clumsily.</td>
<td>Correct stitch position; adequate size bites placed uniform distance apart; perpendicular to seromuscular layer; appropriate tension; adequate surgical knot; needle retrieved safely</td>
</tr>
<tr>
<td>Ensure cinching of suture between bites</td>
<td>Not done</td>
<td>Done inconsistently</td>
<td>Done at all times</td>
</tr>
<tr>
<td><strong>Circular Stapler Technique</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify orogastric tube in stomach and make a gastrotomy in the gastric pouch</td>
<td>Not identified; no entry into gastric lumen; poor relation between grasper and energy source; excessively large or small; penetration of posterior bowel wall; skiving; bleeding</td>
<td>Identified; entry into gastric lumen; appropriate size; more than 1 attempt required</td>
<td>Identified; entry into gastric lumen; appropriate size; no extra movements required</td>
</tr>
<tr>
<td>Pull orogastric tube through the gastric pouch, bring the anvil into the gastric pouch, and cut the tube from anvil</td>
<td>Excessive force applied to the tube; multiple attempts made</td>
<td>Performed correctly; no excessive force; multiple attempts required</td>
<td>Performed correctly; no excessive force applied</td>
</tr>
<tr>
<td>Make an enterotomy in Roux limb</td>
<td>No entry into bowel lumen; poor relation between grasper and energy source; excessively large or small; penetration of posterior bowel wall</td>
<td>Appropriate size with entry into bowel lumen; not placed in appropriate anatomic location</td>
<td>Appropriate size and placement of enterotomy; good relation of grasper and energy source; no extra movements required</td>
</tr>
<tr>
<td>Dilate port site to accept circular stapler</td>
<td>Not done</td>
<td>Inserted after multiple attempts</td>
<td>Performed correctly</td>
</tr>
<tr>
<td>Insert 21/25 mm EEA circular stapler via port site into Roux limb and advance 7 cm</td>
<td>Unclear of how to insert staple device; incorrect stapler; failure to advance the stapler appropriately</td>
<td>Inserts stapler with hesitation; lacks appreciation of the ideal angle for insertion; moderately traumatic</td>
<td>Inserts stapler correctly; smoothly; controlled</td>
</tr>
<tr>
<td>Perforate wall of Roux limb with spike and connect stapler to anvil</td>
<td>Not done</td>
<td>Done correctly after multiple attempts</td>
<td>Done correctly; minimal number of attempts</td>
</tr>
<tr>
<td>Fire stapler</td>
<td>Uncontrolled fire; failure to release staple rings prior to removal</td>
<td>Controlled fire; staple rings released; moderate trauma to anastomosis</td>
<td>Smooth, controlled fire; gentle removal of stapler</td>
</tr>
<tr>
<td>Protect soft tissue around port site against contamination when removing stapler from abdomen</td>
<td>Not done</td>
<td>Done after a delay; prompting required</td>
<td>Done appropriately without prompting</td>
</tr>
</tbody>
</table>
| Remove stapler, replace port | Not done; used excessive force upon removal | Performed correctly; several attempts required | Performed correctly in a smooth motion; no
### Close Roux limb enterotomy using a liner cutting stapler or single layer running suture

| Poorly positioned stitch or stapler. Blindly placed sutures; traumatic use of needle drivers; skiving and poor control of needle; too much or too little tension; poor quality knot | Adequate stitch or stapler position. Adequate closure of enterotomy. Sutures placed at varying distances apart; gathering of bowel edges; additional reinforcement sutures required; stitch pulled out clumsily. | Correct stitch or stapler position. Adequate size bites placed uniform distance apart; appropriate tension; adequate surgical knot; needle retrieved safely |

### Hand-Sewn Technique

<table>
<thead>
<tr>
<th>Approximate Roux limb to gastric pouch end-to-end with a posterior layer of seromuscular running suture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly positioned stitch; blindly placed sutures; traumatic use of needle drivers; skiving and poor control of needle; too much or too little tension; poor quality knot</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Create enterotomy in gastric pouch and Roux limb</th>
</tr>
</thead>
<tbody>
<tr>
<td>No entry into bowel lumen; poor relation between grasper and energy source; excessively large or small; penetration of posterior bowel wall; skiving; bleeding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Place a full thickness posterior layer running suture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly positioned stitch; blindly placed sutures; traumatic use of needle drivers; skiving and poor control of needle; too much or too little tension; poor quality knot</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Place an anterior full thickness running suture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly positioned stitch; blindly placed sutures; traumatic use of needle drivers; skiving and poor control of needle; too much or too little tension; poor quality knot</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Insert a 34/36 F gastric tube / boogie and pass it across the anastomosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not performed</td>
</tr>
</tbody>
</table>

### Closure of potential hernia sites:

#### Closure of JJ Mesenteric Defect

<table>
<thead>
<tr>
<th>Use non-absorbable suture to close the defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly positioned stitch; defect not completely closed; blindly placed sutures; compromised perfusion to the limb; traumatic use of needle drivers; skiving and poor control of needle</td>
</tr>
</tbody>
</table>

#### Closure of Petersen’s Defect

<table>
<thead>
<tr>
<th>Pull transverse colon up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not performed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Place a non-absorbable suture between small bowel mesentery and mesocolon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorly positioned stitch; defect not completely closed; blindly placed sutures; compromised perfusion to the limb; traumatic use of needle drivers; too much or too little tension; poor quality</td>
</tr>
<tr>
<td>Task</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Move small bowel to patient’s left side</td>
</tr>
<tr>
<td><strong>Closure of Transverse Mesocolon Defect</strong></td>
</tr>
<tr>
<td>Use a non-absorbable suture to close a defect between small bowel mesentery and transverse mesocolon</td>
</tr>
<tr>
<td><strong>Testing of GJ anastomosis:</strong></td>
</tr>
<tr>
<td><strong>Blue Dye Technique</strong></td>
</tr>
<tr>
<td>Clamp Roux limb distal to GJ anastomosis</td>
</tr>
<tr>
<td>Fill gastric pouch with 60-120 cc of methylene blue dye</td>
</tr>
<tr>
<td>Suction at left and right corners of anastomosis looking for dye leak</td>
</tr>
<tr>
<td>Request for fluid to be sucked back from gastric pouch</td>
</tr>
<tr>
<td><strong>Air Insufflation Test</strong></td>
</tr>
<tr>
<td>Introduce an endoscope / orogastric tube into gastric pouch</td>
</tr>
<tr>
<td>Instill saline into peritoneal cavity and submerge the gastrojejunostomy in saline</td>
</tr>
<tr>
<td>Occlude Roux limb junostomy distal to GJ anastomosis</td>
</tr>
<tr>
<td>Insufflate gastric pouch with air</td>
</tr>
<tr>
<td>Check for air leak</td>
</tr>
<tr>
<td>Suction saline from peritoneal cavity</td>
</tr>
<tr>
<td><strong>Removal of ports and retractors / fascial closure:</strong></td>
</tr>
<tr>
<td>Unscrew liver retractor and remove in one smooth movement</td>
</tr>
<tr>
<td>Endo-close device to close fascia defects under direct visualization (excluding dilating trocar sites)</td>
</tr>
<tr>
<td>Retract all ports under vision</td>
</tr>
<tr>
<td>Remove intra-abdominal gas</td>
</tr>
</tbody>
</table>
3.4.3 Determination of Feasibility of Use, Validity and Reliability of the BOSATS Scale

3.4.3.1 Reliability

A total of 52 video recordings of LRYGB were reviewed and scored by two independent, trained and blinded raters using the BOSATS scale. The inter-rater reliability coefficient for the BOSATS scale was ICC = 0.954 (p < 0.001). The inter-rater reliability coefficients for individual components of the BOSATS scale ranged from ICC = 0.481 to 0.924 (p < 0.05) (Table 12). The test-retest reliability coefficients for the BOSATS scale were ICC = 0.992 (p < 0.001) and ICC = 0.985 (p < 0.001) for Rater 1 and Rater 2, respectively.

Table 12. Inter-rater reliability coefficients for the entire BOSATS scale and its individual components.

<table>
<thead>
<tr>
<th>Component of BOSATS scale</th>
<th>n</th>
<th>ICC</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Score</td>
<td>52</td>
<td>0.954</td>
<td>0.922 - 0.974</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Access and Port Insertion</td>
<td>23</td>
<td>0.924</td>
<td>0.823 - 0.967</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Placement of Liver Retractor</td>
<td>34</td>
<td>0.721</td>
<td>0.508 - 0.850</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Creation of the Roux Limb</td>
<td>27</td>
<td>0.481</td>
<td>0.123 - 0.723</td>
<td>0.006</td>
</tr>
<tr>
<td>Creation of JJ anastomosis</td>
<td>29</td>
<td>0.813</td>
<td>0.640 - 0.908</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Dissection of the Phreno-Esophageal Ligament</td>
<td>34</td>
<td>0.485</td>
<td>0.189 - 0.703</td>
<td>0.001</td>
</tr>
<tr>
<td>Creation of gastric pouch</td>
<td>50</td>
<td>0.717</td>
<td>0.549 - 0.829</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Positioning of the Roux limb</td>
<td>46</td>
<td>0.830</td>
<td>0.713 - 0.902</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Creation of GJ anastomosis (Linear Stapler)</td>
<td>26</td>
<td>0.647</td>
<td>0.348 - 0.825</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Creation of GJ anastomosis (Circular Stapler)</td>
<td>10</td>
<td>0.639</td>
<td>0.113 - 0.893</td>
<td>0.013</td>
</tr>
<tr>
<td>Creation of GJ anastomosis (Hand Sewn)</td>
<td>15</td>
<td>0.831</td>
<td>0.569 - 0.940</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Closure of JJ mesenteric defect</td>
<td>21</td>
<td>0.704</td>
<td>0.401 - 0.868</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Testing of GJ anastomosis (Air Insufflation)</td>
<td>27</td>
<td>0.877</td>
<td>0.750 - 0.942</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Removal of liver retractor</td>
<td>41</td>
<td>0.726</td>
<td>0.540 - 0.844</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Data reported as n, number of cases analyzed.
BOSATS, Bariatric Objective Structured Assessment of Technical Skill; GJ, gastrojejunostomy; ICC, intraclass correlation coefficient; JJ, jejunooejunostomy.
3.4.3.2 Feasibility of use

The average time to evaluate a 120-minute video recording of an uncomplicated LRYGB operation using the BOSATS scale was 30-40 min.

3.4.3.3 Construct validity

Using case-volume criteria for group assignments, significant differences between novice and experienced surgeon groups were seen for the following components of the BOSATS scale: creation of JJ, linear stapled GJ, circular stapled GJ and hand sewn GJ (Table 13, Figure 15). Novice surgeon group had a lower median score than experienced surgeon group for creation of gastric pouch, albeit statistical significance was not reached.

Using OSATS GRS criteria for group assignments, significant differences between novice and experienced surgeon groups were seen for the following components of the BOSATS scale: creation of JJ, creation of gastric pouch, linear stapled GJ and hand sewn GJ (Table 14, Figure 16). All circular stapled GJ videos were grouped into the experienced surgeon group using OSATS GRS criteria, prohibiting calculation of construct validity for this component of the BOSATS scale. Surgeons that were grouped into the novice group using case-volume criteria for the circular stapled GJ were minimally invasive surgery fellows who had performed fewer than 10 LRYGB operations ‘skin-to-skin’. These minimally invasive surgery fellows have likely performed other advanced laparoscopic procedures and developed the necessary skills to move their overall technical proficiency closer to the level of an experienced surgeon (score > 28 on the OSATS GRS), thereby they were grouped into an experienced surgeon group.

Table 13. Comparison of scores for different components of the LRYGB as assessed by the BOSATS scale using case-volume criteria for definition of experienced and novice surgeons.

<table>
<thead>
<tr>
<th>Component</th>
<th>Observations, n</th>
<th>BOSATS score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Novice</td>
</tr>
<tr>
<td>Jejunoojejunostomy</td>
<td>22</td>
<td>24.5 (21–27)*</td>
</tr>
<tr>
<td>Jejunoojejunostomy</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Gastric pouch</td>
<td>32</td>
<td>36.5 (31.5–38.5)</td>
</tr>
<tr>
<td>Linear stapled</td>
<td>14</td>
<td>25.5 (22–30)*</td>
</tr>
<tr>
<td>Gastrojejunostomy</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Circular stapled</td>
<td>10</td>
<td>40 (37–41)*</td>
</tr>
<tr>
<td>Gastrojejunostomy</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Hand-sewn</td>
<td>14</td>
<td>15 (13–16)*</td>
</tr>
<tr>
<td>Gastrojejunostomy</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Scores reported as median (interquartile range). *Significant differences between groups (p < 0.025).
BOSATS, Bariatric Objective Structural Assessment of Technical Skill Scale.
Horizontal bars, boxes, and whiskers represent the median, interquartile range, and range, respectively. Points outside the interquartile range represent outliers within the group (data points that are beyond the quartile by one-and-a-half interquartile range). BOSATS, Bariatric Objective Structured Assessment of Technical Skill; GJ, gastrojejunostomy; JJ, jejunojejunostomy.

Figure 15. BOSATS scale scores for novice and experienced surgeon groups. Novice and experienced surgeons defined using case-volume criteria.
Table 14. Comparison of scores for different components of the LRYGB as assessed by the BOSATS scale using OSATS GRS criteria for the definition of experienced (OSATS ≥ 28) and novice (OSATS < 28) surgeons.

<table>
<thead>
<tr>
<th>Component</th>
<th>Observations, n</th>
<th>Novice</th>
<th>Experienced</th>
<th>Novice</th>
<th>Experienced</th>
<th>p Value</th>
<th>Maximum possible score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jejunojejunostomy</td>
<td>22</td>
<td>26.5 (24-33)</td>
<td>30.5 (27-32)</td>
<td>0.007</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastric pouch</td>
<td>20</td>
<td>34.5 (29.5-38)</td>
<td>38 (35-41)</td>
<td>0.008</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear stapled gastrojejunostomy</td>
<td>16</td>
<td>27 (23-29.5)</td>
<td>31.5 (29-32)</td>
<td>&lt;0.0001</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circular stapled gastrojejunostomy</td>
<td>0</td>
<td>-</td>
<td>41.5 (40-44.5)</td>
<td>-</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand sewn gastrojejunostomy</td>
<td>14</td>
<td>15 (13-16)</td>
<td>22.5 (18-25)</td>
<td>0.0001</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experienced surgeons defined as Objective Structured Assessment of Technical Skill ≥ 28; novice surgeons as < 28. Scores are reported as median (interquartile range). BOSATS, Bariatric Objective Structured Assessment of Technical Skill.

Surgeons who were grouped into the novice group using case-volume criteria for the circular stapled gastrojejunostomy were minimally invasive surgery fellows with an overall technical proficiency of Objective Structured Assessment of Technical Skill ≥ 28, thereby they were grouped into an experienced surgeon group.
Horizontal bars, boxes, and whiskers represent the median, interquartile range, and range, respectively.

BOSATS, Bariatric Objective Structured Assessment of Technical Skill; GJ, gastrojejunostomy; JJ, jejunoojejunostomy.

Surgeons who were grouped into the novice group using case-volume criteria for the circular stapled gastrojejunostomy were minimally invasive surgery fellows with an overall technical proficiency of Objective Structured Assessment of Technical Skill ≥ 28, thereby they were grouped into an experienced surgeon group.

Figure 16. BOSATS scale scores for novice and experienced surgeon groups. Novice and experienced surgeons defined using the Objective Structured Assessment of Technical Skills Global Rating Scale criteria.
3.4.3.4 Concurrent validity

Moderate to strong correlations between BOSATS scores and OSATS GRS scores were noted for the following components of laparoscopic gastric bypass: creation of JJ, creation of gastric pouch, linear stapled GJ and hand-sewn GJ (Table 15, Figure 17).

Table 15. Spearman's rank correlation coefficients for comparisons between the BOSATS and the OSATS GRS scores for different components of the LRYGB.

<table>
<thead>
<tr>
<th>Component of the LRYGB procedure</th>
<th>N</th>
<th>Spearman’s Rho</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creation of jejunojejunostomy</td>
<td>27</td>
<td>0.59</td>
<td>0.001</td>
</tr>
<tr>
<td>Creation of gastric pouch</td>
<td>49</td>
<td>0.48</td>
<td>0.001</td>
</tr>
<tr>
<td>Creation of linear stapled gastrogejunostomy</td>
<td>25</td>
<td>0.70</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Creation of circular stapled gastrojejunoostomy</td>
<td>10</td>
<td>0.41</td>
<td>0.245</td>
</tr>
<tr>
<td>Creation of hand sewn gastrojejunoostomy</td>
<td>14</td>
<td>0.96</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

n, number of cases reviewed.
BOSATS, Bariatric Objective Structured Assessment of Technical Skill; OSATS GRS, Objective Structured Assessment of Technical Skills Global Rating Scale.

Figure 17. Spearman’s rank correlation for the BOSATS and the OSATS GRS scores for the creation of hand sewn GJ anastomosis.
3.5 Discussion

A scale for objective assessment of operative performance in LRYGB (Bariatric Objective Structured Assessment of Technical Skill - BOSATS) was developed and validated using a systematic and evidence-based approach. Firstly, a hierarchical task analysis of various surgical approaches to LRYGB was conducted to generate a list of all the steps required to complete this procedure. Secondly, this list was presented - in the form of an online Delphi questionnaire - to a panel of international experts in bariatric surgery with a goal of identifying steps that were essential for inclusion into the BOSATS scale. Anchoring descriptors for levels of surgical proficiency were then assigned to each step in the scale. Lastly, feasibility of use, reliability and validity of the BOSATS scale were established using blinded and independent review of prospectively recorded LRYGB procedures.

The BOSATS scale has several strengths. It is the only, objective procedure-specific assessment scale specifically developed and validated for use in LRYGB. The items for this scale were selected using a Delphi method and consensus of an international panel of experts in bariatric surgery – providing evidence for content validity. The BOSATS scale was intentionally designed to address multiple approaches to LRYGB, thereby increasing its transferability between surgeons and institutions.

Procedure-specific assessment scales, such as the scale developed in the present study, are distinctly different from task-specific checklists and global rating scales. Task-specific checklists, such as those for laparoscopic sigmoid colectomy (Sarker et al., 2010), Nissen fundoplication (Cao et al., 1999), cholecystectomy, and inguinal hernia repair (Cao et al., 1999), deconstruct the operation into distinct tasks and mandate a surgeon to perform those tasks in a predefined order. Failure to follow this predefined order often results in a low score on the checklist. Checklists have been shown to provide trainees with structured formative feedback; however, their rigid nature has led to the finding of poor validity and reliability when used by proficient surgeons (Regehr et al., 1998). Global rating scales, in contrast, examine global parameters of operative performance. Global rating scales tend to be more reliable than task-specific checklists (J. A. Martin et al., 1997; Regehr et al., 1998); however, global rating scales are not very useful for providing specific feedback to the learner (Larsen et al., 2008). Procedure-specific assessment scales form the middle ground between task-specific checklists and global rating scales (Palter & Grantcharov, 2012b). Procedure-specific checklists are less rigid than task-specific checklists, while still providing specific formative feedback to the learner.
The Delphi method used in this study had several strengths. It offered the opportunity to conduct the questionnaire online, thereby improving feasibility and lessening costs (Fink et al., 1984). This approach allowed for recruitment of participants from various geographical locations (Graham et al., 2003), while the anonymous nature of the Delphi method made it challenging for a single influential participant to have a disproportionate impact on the outcome of the questionnaire (Graham et al., 2003). The response rates for the 1st and 2nd round of the questionnaire were greater than the response rates of 54-61% commonly reported in health professional literature (Burns et al., 2008; Palter, MacRae, et al., 2011; Sierles, 2003), which was likely a reflection of participants’ interest in the presented topic. Despite excellent response rates, however, two rounds of the online Delphi questionnaire were required to achieve consensus among the panel members on the steps of LRYGB that were of importance for inclusion into the BOSATS scale (Table 10). This lack of initial consensus speaks to the diversity of opinions among international experts on the steps of LRYGB of importance for inclusion into the BOSATS scale. The steps associated with a low alpha score suggested a greater degree of disagreement between panel members regarding inclusion of that step into the BOSATS scale. Studies by Palter & Grantcharov (Palter, Graafland, Schijven, & Grantcharov, 2012; Palter, MacRae, et al., 2011), which used similar methodologies to develop an assessment scale for laparoscopic colorectal surgery, reported the need for more than one round of online Delphi questionnaire to reach consensus.

The BOSATS scale demonstrated excellent test-retest reliability, very high inter-rater reliability (ICC > 0.9) for total score and moderate to high (ICC 0.6-0.9) reliability for component scores. Inter-rater reliability is a psychometric property of a scale that refers to the extent to which ratings of the same performance by different observers are similar (Vassiliou et al., 2005). Deliberate calibration and training of raters was likely a contributing factor to these results. In fact, the BOSATS scale showed greater inter-rater reliability than both GOALS and OSATS GRS (J. A. Martin et al., 1997) scales. The test-retest reliability was also very high (ICC = 0.98), more than satisfying the predefined cut-off of 0.8 that has been suggested to be acceptable (Gallagher, Ritter, & Satava, 2003).

The feasibility of use for the BOSATS scale was demonstrated by noting a relatively small time commitment of 30-40 min for assessment of one LRYGB video recording. It could be argued that 40 minutes is a long time to spend assessing one LRYGB case, however, this time investment to objectively identify areas of weakness and to guide deliberate practice outside the
operating room is expected to shorten the learning curves and result in cost savings in the operating room (Orzech et al., 2012). Most of the minimally invasive operating rooms in academic hospitals have the necessary equipment to permit video recording of cases, thereby making this form of assessment feasible and easy to use.

Construct validity for the components of the BOSATS scale was demonstrated by comparing the scores of novice and experienced surgeons for different components of the LRYGB operation. Construct validity refers to the degree to which a score can be interpreted as representing the intended underlying construct (Cook & Beckman, 2006). Comparison of total BOSATS scores, rather than component scores, between novice and experienced surgeons was not possible, because novice surgeons often learned this complex laparoscopic operation one component at a time (Zevin, Aggarwal, et al., 2012). A trainee would likely achieve proficiency in the JJ prior to being allowed to attempt the creation of the gastric pouch. Thus, a novice surgeon for the gastric pouch component would likely be an experienced surgeon for the JJ component. In regards to the definitions of novice and experienced surgeons, Ericsson (Ericsson, 2007) noted that years of experience, academic standing or rank, and specialty board certification have often been equated to ‘expertise’ and suggested that a better approach may be to determine expertise through reproducibly superior performance (improved treatment outcomes, diagnoses, etc.). Taking that recommendation into account, this study used case-volumes (< 10 and > 100 cases) and objective scoring metrics (OSATS GRS score < 28 and >/= 28) for definitions of novice and experienced surgeons. As a result, the number of novice and experienced surgeons in each group varied depending on the definition used. Using the OSATS GRS score, five novice surgeons that performed a circular stapled GJ were re-classified into the experienced surgeon group (Table 6 and 7). These surgeons were minimally invasive fellows who had performed fewer than 10 LRYGB operations ‘skin-to-skin’; however, they would have likely performed other advanced laparoscopic procedures and developed the necessary skills to move their overall technical proficiency closer to the level of an experienced surgeon. Using case-volume criteria, significant differences between experienced and novice groups were seen for four out of the five components analyzed (Table 6). Creation of the gastric pouch is also often the last component of the operation that is taught to a trainee. Thus, a surgeon would often be proficient in other advanced laparoscopic techniques by the time he or she begins to create the pouch. This statement is supported by the finding of greater number of surgeons in the experienced group versus novice group when surgeons are reclassified using OSATS GRS (Tables 6 and 7). Consequently, differences between novice and experienced surgeons for the gastric pouch
component were small, and a larger sample size would be required to demonstrate a potential statistically significant difference.

Concurrent validity is another measure of validity for the BOSATS scale. It is defined as “an evaluation in which the relationship between the test scores and the scores on another instrument purporting to measure the same construct are related” (Gallagher et al., 2003). BOSATS scores were correlated with OSATS GRS scores for various components of the operation. Moderate to high correlations were noted for creation of JJ (rho = 0.59), gastric pouch (rho = 0.48), linear (rho = 0.70) and hand sewn GJ (rho = 0.96) suggesting the both scales are measuring components of the same construct – operative performance. No significant correlation was noted for the circular GJ component. This lack of correlation was a result of the relative homogeneity of surgeons who performed circular stapled anastomoses - all surgeons scored ≥ 28 on the OSATS GRS scale.

This study had a number of limitations. First, missing data points for the Delphi questionnaire were generated when panel members did not rate all the items, however, regardless of how the missing data points were analyzed (replaced by mean, mode or neutral “3”), Cronbach’s alpha remained above the predetermined consensus cut-off of 0.80. Second, a modification to the Delphi methodology was made for round 2 of the Delphi questionnaire for feasibility reasons. Greater number of panel members was recruited for (n = 19) and responded to round 2 of the questionnaire (n = 17) compared to round 1 (n = 12). Within the classical Delphi methodology, same participants complete each of the rounds of the Delphi process. In this study, the degree of heterogeneity of opinion regarding steps of the LRYGB for inclusion into the BOSATS scale and 12 individuals within the Delphi panel resulted in no consensus in round 1 of the questionnaire. To mitigate this problem, seven additional individuals were asked to participate as members of the Delphi panel for round 2 of the questionnaire. The increase in the Delphi panel membership for round 2, which is a modification to the classical Delphi methodology, was done for feasibility reasons. They may have introduced bias within our results; however, individuals who were invited to participate in round 2 met the same inclusion criteria for the Delphi panel membership as individuals in round 1.

Third, it was also not possible to demonstrate construct validity for all components of the BOSATS scale (i.e. ‘patient positioning’, ‘access of peritoneal cavity with optical viewing trocar without prior abdominal insufflation’, ‘triple stapling technique for creation of
jejunojejunostomy’, etc.), because some steps of LRYGB procedure were not recorded with the intraabdominal camera view (e.g. ‘patient positioning’) and surgeons who participated in this study did not utilize all possible approaches to the LRYGB. Future projects should focus on demonstration of construct and concurrent validity of the BOSATS scale for those additional components.

Fourth, this study excluded a group of surgeons with an intermediate level of experience (10 – 100 cases). This group was excluded for feasibility reasons, because there are only two groups of surgeons at the University of Toronto, those with minimal experience in advanced laparoscopic bariatric surgery (surgery residents and minimally invasive surgery fellows at the start of their fellowship), and those with extensive experience in advanced laparoscopic bariatric surgery (fellowship trained attending surgeons). There are fewer than 5 individuals at a given time with an intermediate experience in advanced laparoscopic bariatric surgery (10-100 cases), because this group represents individuals at the end of their minimally invasive surgery fellowships.

Fifth, there were only two raters in this study. Both raters were trained in the use of BOSATS scale and were blinded to the level of experience of the study participants. Future studies with greater number of rates, and non-trained rates are required to confirm the reliability results achieved in this study.

Lastly, it is important to emphasize that the BOSATS scale was designed and validated for use in formative, rather than summative assessment. At present, it is intended for use in identification of specific areas of weakness, provision of specific targeted feedback, and facilitation of deliberate practice. Additional studies are required to generate specific proficiency cut-off values for use in high-stakes assessment, certification and re-certification.
3.6 Conclusion

The Bariatric Objective Structured Assessment of Technical Skill scale appears to be feasible to use, reliable and valid instrument for objective assessment of operative performance in LRYGB. The list of steps generated by the Delphi expert panel created a “training itinerary”, which can be used to guide the development of a comprehensive training curriculum for laparoscopic bariatric surgery. Implementation of the BOSATS scale in such a curriculum has the potential to provide trainees with objective structured feedback, facilitate deliberate practice and shorten learning curves in the operating room.
Chapter 4

Development and Validation of a Comprehensive Simulation-Enhanced Training Curriculum for a Complex Minimally Invasive Operation: A Randomized Controlled Trial
4 Development and Validation of a Comprehensive Simulation-Enhanced Training Curriculum for a Complex Minimally Invasive Operation: A Randomized Controlled Trial

4.1 Abstract

Context: Simulation-enhanced training has been shown to improve technical and non-technical skills in surgery. The effectiveness of a comprehensive ex-vivo curriculum in comparison to conventional training in advanced minimally invasive procedures is still unclear.

Objectives: (1) To develop a comprehensive, simulation-enhanced training (SET) curriculum for laparoscopic bariatric surgery and (2) to compare its effectiveness to conventional surgery training (CST).

Design, Setting and Participants: A prospective, single-blinded randomized controlled trial conducted at the University of Toronto (May 2012 – March 2013) with 20 postgraduate year (PGY) 3 and 4, and 12 final year (PGY 5) general surgery residents.

Intervention: Twenty PGY 3 and 4 residents were randomly assigned to the SET (n = 10) or CST (n = 10) group. The SET group completed a comprehensive curriculum (cognitive, technical, and non-technical components), while the CST group underwent conventional training. Four months after study enrolment, each participant completed a laparoscopic jejunojejunostomy (JJ) in a live anesthetized porcine model. Participants continued to practice in that model until a minimum level of clinical proficiency was achieved.

Main Outcome Measures: We quantified operative skill on a live porcine JJ using the Bariatric Objective Structured Assessment of Technical Skills (BOSATS) scale, non-technical skill in a simulated operating room (OR) crisis scenario using the Non-Technical Skills for Surgeons (NOTSS) scale; operative skill on a laparoscopic JJ in the OR using the BOSATS scale; and knowledge using a 25-question test.

Results: Twenty out of the 26 eligible participants agreed to participate. The SET group demonstrated superior operative skill in a live porcine model (BOSATS: 56.4(11.5) versus 46.0(10.6), P < 0.05) and superior non-technical skills (NOTSS: 40.8(4.2) versus 31.6(8.7), P <
The SET group required fewer cases in a live porcine model and had a greater number of participants with a minimum level of clinical proficiency by their 2nd case (1/7 versus 4/5, \( P = 0.03 \)). The SET and CST group had equivalent knowledge (13.9(1.8) versus 12.8(2.9), \( P = 0.32 \)) and operative skills in the OR (BOSATS: 64.1(10.8) versus 61.1(8.8), \( P = 0.53 \)). In comparison to the PGY 5 group, the SET group had equivalent operative skill in the OR (BOSATS: 64.1(10.8) versus 67.3(9.6), \( P = 0.53 \)), superior non-technical skills (NOTSS: 40.8(4.2) versus 31.3(6.2), \( P < 0.01 \)) and less knowledge (12.8(2.9) versus 15.8(2.1), \( P = 0.02 \)).

**Conclusions:** Participation in a comprehensive curriculum resulted in superior training outcomes when compared to conventional surgery training. Implementation of this curriculum may offer standardization of bariatric surgery training and ensure that comprehensive proficiency milestones are attained prior to exposure to patient care.

**Trial Registration:** www.clinicaltrials.gov Identifier: NCT01610466
4.2 Introduction

Over the past two decades, simulation-enhanced education has gained increased popularity as an adjunct to traditional training in health professions (Cook et al., 2011). Several meta-analyses and systematic reviews have demonstrated an association between simulation training and improvements in knowledge, skills and behaviors of learners, as well as patient-related outcomes (Cook et al., 2012; Cook et al., 2011; Issenberg et al., 2005). In the context of surgery training, simulation-enhanced education has been shown to translate to improvements in knowledge (Swanstom et al., 2006), technical skills (Larsen et al., 2009; Swanstrom et al., 2006), and teamwork and team-performance (Gettman et al., 2009) in the operating room. More importantly, there is emerging evidence that simulation-enhanced training is associated with improved intra-operative and post-operative patient outcomes (Zendejas et al., 2011).

Despite this growing body of evidence supporting the use of simulation in surgery education and the multiple calls from prominent surgical educators to embed simulation training into structured training curricula (Rajesh Aggarwal & Darzi, 2011; Palter & Grantcharov, 2010; Windsor, 2011), minimal progress has been made in the area of simulation training in advanced minimally invasive surgery. The American College of Surgeons and the Association of Program Directors in Surgery developed a Surgical Skills Curriculum for Residents that focuses on basic and intermediate laparoscopic operations and team-based skills ("ACS/APDS Surgical Skills Curriculum for Residents,"). In a research setting, structured simulation-enhanced training curricula have been developed for basic laparoscopic skills (Swanstom et al., 2006), as well as basic (R. Aggarwal et al., 2009b) and intermediate laparoscopic operations (Palter & Grantcharov, 2012a; Zendejas et al., 2011). These curricula concentrate primarily on training of technical skills and fail to address such non-technical skills as teamwork, decision making, situation awareness and leadership. They do not address complex laparoscopic procedures and do not conform to one set of published recommendations for the design, validation, and implementation of simulation-based training curricula in surgery (Zevin, Levy, Satava, & Grantcharov, 2012). A comprehensive, simulation-enhanced training (SET) curriculum for complex laparoscopic surgery was warranted in view of the evidence that poor technical (Cuschieri, 2005) and non-technical skills were associated with increased errors in the operating room (Catchpole, Mishra, Handa, & McCulloch, 2008) and that the technical skills of surgeons correlated strongly with complication rates in complex laparoscopic procedures (Birkmeyer, 2012).
The objectives of this study were (1) to develop a comprehensive, SET curriculum for laparoscopic bariatric surgery, and (2) to compare the effectiveness of the SET curriculum to conventional surgery training (CST).

### 4.3 Hypothesis

We hypothesized that the SET group would demonstrate (1) superior technical skills, (2) superior non-technical skills, and (3) superior knowledge of bariatric surgery in comparison to the CST group.
4.4 Methods

4.4.1 Study Design & Dates

This was a prospective, single-blinded, randomized controlled trial conducted at the University of Toronto between May 2012 and March 2013. The research ethics board of the University of Toronto approved this study. Study flow diagram is presented in Figure 18. The CONSORT diagram is presented in Figure 19.

![Study Flow Diagram]

JJ, jejunojejunostomy; SET, simulation-enhanced training; CST, conventional surgery training; OR, operating room.

Figure 18. Study Flow Diagram.
4.4.2 Participants

Post-graduate year (PGY) 3 and 4 general surgery residents at the University of Toronto were recruited. Participants were randomly allocated to either the SET group (n = 10) or the CST group (n = 10). Twelve final-year (PGY-5) general surgery residents were recruited as a benchmark group to represent trainees who were “ready for independent practice”. All participants signed an informed consent form.

4.4.3 Inclusion and Exclusion Criteria

PGY 3 and 4 general surgery residents who had performed fewer than 10 advanced laparoscopic operations as the primary surgeon (laparoscopic bariatric surgery, laparoscopic low anterior resection, laparoscopic esophagectomy, laparoscopic pancreatectomy, laparoscopic splenectomy,
laparoscopic adrenalectomy) were included. Final-year (PGY-5) general surgery residents were included irrespective of their advanced laparoscopic experience to represent trainees that were deemed “ready for independent practice”.

4.4.4 Baseline Demographic and Clinical Data

Age, sex, year of training, number of basic, intermediate and advanced laparoscopic procedures completed as the primary surgeon, and clinical rotations completed to date were collected at the start of the study. Basic laparoscopic procedures were defined as: diagnostic laparoscopy, laparoscopic appendectomy and laparoscopic cholecystectomy. Intermediate laparoscopic procedures were defined as: laparoscopic Nissen fundoplication, laparoscopic ventral hernia repair and laparoscopic colon resections (excluding rectal resections).

4.4.5 Baseline Assessment of Technical Skill

All participants performed a laparoscopic jejunojejunostomy in a box trainer on cadaveric porcine small bowel for baseline assessment of technical skill. Participant’s performance was recorded through a laparoscopic camera and was assessed by a trained rater (EB), blinded to the identity, training level and group allocation of the participant. The Bariatric Objective Structured Assessment of Technical Skills (BOSATS) scale (Zevin et al., 2013b) was used for assessment of technical skill.

4.4.6 Intervention

The intervention in this study was participation in a comprehensive, SET curriculum. The SET curriculum was composed of cognitive, technical and non-technical components and was created in accordance with a prior consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery (Zevin, Levy, et al., 2012).

4.4.6.1 Cognitive component

Participants were provided reading materials on the topic of laparoscopic bariatric surgery (Ashley, 2012; Townsend, Beauchamp, Evers, & Mattox, 2012) and attended a 2-hour faculty-led seminar addressing key technical steps for laparoscopic Roux-en-Y gastric bypass, laparoscopic sleeve gastrectomy and laparoscopic placement of an adjustable gastric band.
4.4.6.2 Technical component

Participants practiced on a laparoscopic JJ model in a box trainer to a predefined level of proficiency (BOSATS score ≥ 80%). This proficiency level was set based on published BOSATS scores of experienced surgeons for the JJ component of the LRYGB (Zevin et al., 2013a). The laparoscopic JJ model consisted of a laparoscopic tower, light source and camera (Karl Storz – Endoskope, Tuttlingen, Germany), a box-trainer (3-D Med, Franklin, OH, USA), a 2-foot piece of cadaveric porcine small bowel with mesentery, a laparoscopic linear cutting stapler (3.5mm, 60 mm, Ethicon, Cincinnati, OH, USA), laparoscopic instruments, and an energy source for electrocautery dissection (Force FX-CS Generator, Covidien, Mansfield, MA, USA). A member of the study team (BZ) supervised each session, provided objective, real-time feedback and one-on-one instruction, and evaluated participants’ technical skill during the session. Each training session was limited to 1.5 hours, in keeping with a distributed practice schedule for maximum learning effectiveness (Moulton et al., 2006).

4.4.6.3 Non-technical component

The non-technical skills component addressed situation awareness, decision-making, communication and teamwork, and leadership. It was comprised of an interactive, expert-led seminar on non-technical skills in the operating room (60 min), and a simulated intra-operative crisis during laparoscopic surgery, such as tension pneumothorax and anaphylactic reaction to an antibiotic (15 min), followed by a structured debriefing session (15 min). Simulations were conducted in a standardized fashion using a human patient simulator (SimMan 3G, Laerdal Medical, Stavanger, Norway) in a realistic operating room environment. The roles of an anesthesia resident, a medical student, and a scrub nurse were scripted and played by trained members of the study team. A member of the study team (ND) moderated the interactive seminar and the debriefing sessions. All simulations were video recorded.

4.4.7 Conventional Surgery Training (CST)

The CST group continued their progress through the general surgery residency-training program, which included participation in all of the required clinical activities, attendance of 2-hour weekly general surgery didactic teaching sessions and site-specific rounds. Participants in the CST group were permitted to conduct independent study on any topic, including bariatric surgery.
4.4.8 Live Anesthetized Porcine Model

Four months after enrolment into the study, all study participants performed a laparoscopic JJ in a live, anesthetized porcine model. This model was chosen as an intermediate step between training in a box trainer and operating on a patient in the OR to standardize the evaluation of technical skill without potential confounding factors of patient’s sex, body habitus and body mass index. Furthermore, this step was introduced due to ethical concerns of allowing trainees without previous simulation exposure to operate on patients in the operating room. Bariatric surgeons and fellows in minimally invasive surgery supervised all sessions in the live porcine model. Within the context of this study, every participant was required to demonstrate a minimum level of clinical proficiency (BOSATS ≥ 60%) in a live porcine model prior to progressing to the operating room. A minimum level of clinically proficiency was defined as a mean score of 3 out of 5 on the BOSATS scale (Zevin et al., 2013a). Preceptors assessed the technical skill of the participants in real time using the BOSATS scale. The decision to progress to the operating room was based on the real time scores of preceptors. Each laparoscopic JJ on a live, anesthetized porcine model was also video-recorded and subsequently assessed by a trained rater (EB), blinded to the identity, training level and group allocation of the participant. The scores of the blinded rater (EB), rather than the scores of the non-blinded preceptors were used for the analysis.

4.4.9 Primary Outcome

The primary outcome of this study was the technical performance on the laparoscopic JJ in a live, anesthetized porcine model measured using the BOSATS scale.

4.4.10 Secondary Outcomes

4.4.10.1 Technical skill in the operating room

Each participant performed one laparoscopic JJ on a patient in the operating room within 5 months (mean 3.3 months, standard deviation 1.2 months) of completion of training in a live porcine model. All procedures were performed using a standardized operative technique under the supervision of one experienced bariatric surgeon (TG) blinded to the group allocation of the participant. The number and duration of takeovers by the supervising surgeon, as well as patient’s age, body-mass index (BMI), sex and the American Society of Anesthesiology (ASA) classification were recorded. Operative performance was video recorded through the laparoscopic camera and submitted for assessment by a trained rater (EB) blinded to the identity,
training level and group allocation of the participant. The BOSATS scale for the creation of laparoscopic jejunojejunostomy (Figure 20) and the Objective Structured Assessment of Technical Skill Global Rating Scale (OSATS GRS) were used for assessment of operative performance (Figure 4).

**Participant ID: ____________

Video: □ Pretest JJ □ Live Animal JJ, Test Session # ________ □ OR JJ

**BOSATS Scale:**

<table>
<thead>
<tr>
<th>Task / Step</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>n/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure approximately 40-60cm of jejunal distal to the ligament of Treitz</td>
<td>Length not measured</td>
<td>Measured, however individual measurements not of the same size; poor orientation</td>
<td>Measured methodologically; each measurement of the same size; correct orientation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confirm that this part of jejunum will reach the hiatus / gastric pouch</td>
<td>Not confirmed</td>
<td>Confirmed briefly</td>
<td>Confirmed clearly and methodologically</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divide jejunum</td>
<td>Divided with tissue trauma and gross contamination</td>
<td>Bowel divided; jaws of stapler not perpendicular to bowel; possibility of mesenteric trauma; need for re-resection</td>
<td>Bowel divided at correct angle; good viability of ends; no trauma to mesentery</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Creation of JJ anastomosis:**

| Measure 75-150 cm length of Roux limb              | Length not measured | Measured, however individual measurements not of the same size; poor orientation; grabs onto mesentery | Measured methodologically; each measurement of the same size; correct orientation |         |         |     |
| Bring distal end of BP limb and position side to side to Roux limb | Incorrect orientation of BP limb to Roux limb | Correct orientation, although required multiple attempts with extra movements | BP limb positioned correctly in relation to Roux limb; no extra movements; no tension |         |         |     |
| Create enterotomies in BP and Roux limbs          | Poor relation between grasper and energy source; excessively large or small; penetration of posterior bowel wall | Appropriate size enterotomy; not placed in antimesenteric position | Appropriately sized and placed enterotomies; no extra movements. Good relation of grasper and energy source |         |         |     |
| Insert the limbs of linear cutting stapler into the enterotomies in Roux and BP limbs | Unclear of how to insert the staple device. Drives staple jaws blindly into BP and Roux limb | Inserts the stapler with hesitation and lacks appreciation of the ideal angle for insertion | Inserts staple jaws with ease; controlled manner; correct angle |         |         |     |
| Ensure both limbs are symmetrical and stapler on anti-mesenteric border | Does not ensure limb symmetry and anti-mesenteric position prior to closure of jaws | Limbs either non-symmetrical or not on anti-mesenteric border prior to closure of jaws | Correctly ensures symmetry and anti-mesenteric position prior to closure of the jaws |         |         |     |
| Fire stapler                                      | Uncontrolled fire with excessive pull on the bowel and widening of enterotomies | Controlled fire; some slippage of bowel from jaws | Smooth, controlled fire; no widening of enterotomies |         |         |     |
| Close the created enterotomy with a simple running suture or a linear cutting stapler | Poorly positioned stitch or stapler. Blindly placed sutures; traumatic use of needle drivers; skiving and poor control of needle; too much or too little tension; poor quality knot | Adequate stitch or stapler position. Adequate closure of enterotomy. Sutures placed at varying distances apart; gathering of bowel edges; additional reinforcement sutures required; stitch pulled out clumsily | Correct stitch or stapler position. Adequate size bites placed uniform distance apart; perpendicular to seromuscular layer; appropriate tension; adequate surgical knot; needle retrieved safely |         |         |     |

4.4.10.2 Performance in a simulated crisis scenario

Each participant completed a 15-minute, simulated intra-operative crisis scenario (intra-operative tension pneumothorax), which differed from the practice scenario for the SET group (intra-operative anaphylactic shock). The same, standardized configuration of high fidelity simulations was used as in the practice session for the SET group. All simulations were video-recorded and evaluated by two independent observers (EB and ND) after the completion of the data collection using the previously validated NOTSS rating scale (Figure 7) (Yule et al., 2008). One of the observers (EB) was blinded to the level of training and group allocation of the participants. The NOTSS scores for each category were calculated by summing the element scores (range of scores from 4 to 16), and the total score was calculated by summing all of the element scores (range of scores from 12 to 48).

4.4.10.3 Knowledge

A score on a 25-question knowledge test was a secondary outcome in this study. This non-validated test was constructed by a member of the study team (BZ), and was designed to evaluate the participant’s understanding of indications, contraindications, complications and technical aspects of laparoscopic bariatric surgery. This test was piloted with PGY3 surgery residents, minimally invasive surgery fellows, and attending surgeons. Significant differences were demonstrated for the test scores between the groups.

4.4.11 Sample Size

This study used a 25-question knowledge test, as well as the BOSATS, OSATS GRS and NOTSS scales for assessment. Using effect sizes reported in prior studies (Crossley, Marriott, Purdie, & Beard, 2011; Lee, Mucksavage, Canales, McDougall, & Lin, 2012; Palter & Grantcharov, 2012a; Zevin et al., 2013b), an alpha of 0.05 in a two-sided Student’s t-test and a power of 0.80, a minimum of 10 participants was required for each study group.

4.4.12 Randomization

The allocation sequence was generated using unrestricted randomization with a non-transparent closed envelope technique. Allocation of all participants to study groups was completed prior to the start of the intervention and data collection.
4.4.13 Confounding Variables

The numbers of basic, intermediate and advanced laparoscopic cases and types of clinical rotations completed during the study period were recorded.

4.4.14 The PGY 5 Group

Twelve final-year (PGY-5) general surgery residents were recruited as a benchmark group to represent trainees who were “ready for independent practice”. Performance on post-intervention assessment (Figure 18) was assessed using the same measures as for participants in the SET and CST groups.

4.4.15 Statistical Analyses

Descriptive statistics were calculated. Student’s t-test was used for between-group comparisons of continuous variables. The Fisher exact test was used for between-group comparisons of categorical variables. A paired t-test was used for within-group comparisons of continuous variables. The NOTSS scores of two observers for crisis scenarios were pooled and analyzed using a t-test with Bonferroni correction for repeated measures (statistical significance was set to $\alpha = 0.025$). Inter-rater agreement was calculated using the intraclass correlation coefficient (2-way mixed effects model, absolute agreement, average measures). Statistical significance was set to $\alpha = 0.05$. Data are presented as means (SD) unless stated otherwise. All statistical analyses were performed using STATA version 12.1.
4.5 Results

A total of 20 out of 26 eligible participants agreed to participate in this study. Baseline demographic characteristics were equivalent between the SET and the CST group, albeit the SET group had participated in significantly fewer bariatric rotations (Table 16). The baseline technical skills were equivalent between the SET and the CST group (Table 17). The number of basic, intermediate, advanced laparoscopic procedures performed during the duration of the study was equivalent between the SET and the CST group (Table 18).

4.5.1 Primary Outcome: Technical Skill in a Live Anesthetized Porcine Model

The SET group achieved a greater technical skill score on a laparoscopic JJ in a live porcine model (p < 0.05) with a large effect size (Cohen’s d = 0.95) (Table 17, Figure 21). The SET group showed within-group improvement in technical skill from baseline assessment in a box trainer to assessment in a live porcine model [BOSATS: 39.0(15.1) versus 56.4(11.5), p < 0.01] with a large effect size (d = 1.27) (Figure 25). The CST group did not show any within-group improvement [BOSATS: 43.5(13.4) versus 46.0(10.6), p = 0.57] (Figure 24). The SET group practiced an average of 2.1(0.8) hours in a box trainer and performed an average of 2.3(0.7) laparoscopic jejunojejunostomies in a box trainer.

Using a predefined, minimum level of clinical proficiency (BOSATS > 60%), the SET group required fewer cases in a live porcine model than the CST group [1.6(0.7) versus 2.2(1.1)]. Five out of 10 participants in the SET group, and 3 out of 10 participants in the CST group achieved a minimum level of clinical proficiency on the 1st laparoscopic jejunojejunostomy and thus were allowed to progress to the operating room. Greater number of remaining participants in the SET group versus the CST group achieved a minimum level of clinical proficiency by their 2nd laparoscopic JJ case [4 out of 5 versus 1 out of 7, P < 0.04] (Table 17). The required time to achieve a predefined, minimum level of clinical proficiency was 1.6(0.7) hours for the SET group, and 2.5(1.8) hours for the CST group.

Blinded rater scores were less than the preceptor scores for each practice session [JJ # 1 (n=20): 51.2(12.0) versus 71.3(16.0), p < 0.01; JJ # 2 (n=12): 50.8(12.2) versus 69.4(14.1), p < 0.01; JJ # 3 (n=3): 51.9(4.6) versus 75(6.6), p < 0.05].
Table 16. Demographic data and laparoscopic experience of the study participants.

<table>
<thead>
<tr>
<th></th>
<th>CST (n = 10)</th>
<th>SET (n = 10)</th>
<th>PGY 5 (n = 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (Male: Female)</td>
<td>5 : 5</td>
<td>5 : 5</td>
<td>7 : 5</td>
</tr>
<tr>
<td>Age (years)</td>
<td>29.8 (3.2)</td>
<td>29.7 (2.4)</td>
<td>30.8 (1.2)</td>
</tr>
<tr>
<td>Hand Dominance (Right : Left)</td>
<td>10 : 0</td>
<td>10 : 0</td>
<td>11 : 1</td>
</tr>
<tr>
<td>Level of Training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGY 3</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>PGY 3 (research)</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>PGY 4</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>PGY 5</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>No. Basic Laparoscopic Cases$^\dagger$</td>
<td>71 (86)</td>
<td>24 (24)</td>
<td>293 (182)</td>
</tr>
<tr>
<td>No. Intermediate Laparoscopic Cases$^\dagger$</td>
<td>8 (13)</td>
<td>1 (1)</td>
<td>107 (97)</td>
</tr>
<tr>
<td>No. Advanced Laparoscopic Cases$^\dagger$</td>
<td>1 (1)</td>
<td>1 (2)</td>
<td>16 (22)</td>
</tr>
<tr>
<td>No. Laparoscopic Bariatric Cases$^\dagger$</td>
<td>0 (1)</td>
<td>1 (2)</td>
<td>42 (29)</td>
</tr>
<tr>
<td>Participation in a Bariatric Rotation* (Yes : No)</td>
<td>7 : 3</td>
<td>2 : 8</td>
<td>9 : 3</td>
</tr>
<tr>
<td>Bariatric Rotation (months)$^\dagger$</td>
<td>1.6 (1.3)</td>
<td>0.6 (1.3)</td>
<td>4.3 (1.4)</td>
</tr>
<tr>
<td>MIS Rotation (Yes : No)</td>
<td>7 : 3</td>
<td>3 : 7</td>
<td>10 : 2</td>
</tr>
<tr>
<td>MIS Rotation (months)$^\dagger$</td>
<td>1.8 (1.5)</td>
<td>0.8 (1.3)</td>
<td>4.7 (2.9)</td>
</tr>
<tr>
<td>Interest in MIS / Bariatric Surgery (Yes : No : Unsure)</td>
<td>2 : 2 : 6</td>
<td>2 : 3 : 5</td>
<td>8 : 4</td>
</tr>
</tbody>
</table>

CST – conventional surgery training; SET – simulation-enhanced training; SD – standard deviation; MIS – minimally invasive surgery rotation (participation in intermediate and advanced laparoscopic cases); No. – number.

*Statistically significant difference between CST and SET groups: Participation in bariatric rotation (Fisher exact test, p = 0.035).

$^\dagger$ = mean (SD)
Table 17. Cognitive, technical and non-technical performance at baseline and post-training for the CST, SET and PGY 5 groups.

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>CST (n = 10)</th>
<th>SET (n = 10)</th>
<th>PGY 5 (n = 9)</th>
<th>Effect Size (d)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline Technical Skill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laparoscopic JJ in box trainer (BOSATS score)</td>
<td>43.5 (13.4)</td>
<td>39.0 (15.1)</td>
<td>-</td>
<td>0.49</td>
<td>-</td>
</tr>
<tr>
<td><strong>Post-Intervention Technical Skill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laparoscopic JJ in live porcine model (BOSATS score)</td>
<td>46.0 (10.6)</td>
<td>56.4 (11.5)</td>
<td>&lt; 0.05</td>
<td>0.95</td>
<td>-</td>
</tr>
<tr>
<td>Laparoscopic JJ in live OR*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOSATS score</td>
<td>61.1(8.8)</td>
<td>64.1(10.8)</td>
<td>-</td>
<td>67.3(9.6)</td>
<td>0.53</td>
</tr>
<tr>
<td>OSATS score</td>
<td>16.7(2.8)</td>
<td>18.9(3.9)</td>
<td>0.19</td>
<td>-</td>
<td>23.6(5.7)</td>
</tr>
<tr>
<td>Proportion of operative steps completed without takeover</td>
<td>0.67(0.26)</td>
<td>0.86(0.16)</td>
<td>0.10</td>
<td>-</td>
<td>0.86(0.14)</td>
</tr>
<tr>
<td><strong>Proportion of participants achieving a minimum level of clinical proficiency on a live porcine model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; attempt</td>
<td>0.63</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; attempt</td>
<td>0.63</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Post-Intervention Non-Technical Skill (NOTSS score)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation Awareness</td>
<td>8.6 (2.7)</td>
<td>10.8 (1.3)</td>
<td>&lt; 0.01</td>
<td>1.02</td>
<td>9.3 (1.4)</td>
</tr>
<tr>
<td>Decision Making</td>
<td>8.2 (2.8)</td>
<td>10.5 (1.5)</td>
<td>&lt; 0.01</td>
<td>1.05</td>
<td>7.4 (1.3)</td>
</tr>
<tr>
<td>Communication and Teamwork</td>
<td>6.9 (1.4)</td>
<td>9.2 (1.4)</td>
<td>&lt; 0.01</td>
<td>1.36</td>
<td>7.2 (2.2)</td>
</tr>
<tr>
<td>Leadership</td>
<td>8.0 (2.4)</td>
<td>10.3 (1.5)</td>
<td>&lt; 0.01</td>
<td>1.17</td>
<td>8.3 (1.5)</td>
</tr>
<tr>
<td>Total</td>
<td>31.6 (8.7)</td>
<td>40.8 (4.2)</td>
<td>&lt; 0.01</td>
<td>1.34</td>
<td>31.3 (6.2)</td>
</tr>
<tr>
<td><strong>Post-Intervention Knowledge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; attempt</td>
<td>13.9 (1.8)</td>
<td>12.8 (2.9)</td>
<td>0.32</td>
<td>-</td>
<td>15.8 (2.1)</td>
</tr>
</tbody>
</table>

Significance for BOSATS and knowledge test was set to p < 0.05; NOTSS scores to p < 0.025.

n – number of participants

* n = 8 for the SET group

P<sup>§</sup> - Comparison between CST and SET groups

P<sup>¶</sup> - Comparison between SET and PGY 5 group

P<sup>♭</sup> - Comparison between CST and PGY 5 group.

Effect sizes reported as Cohen’s d. Highlighted cell represents a statistically significant result.
Table 18. Laparoscopic experience of the study participants during the course of the study.

<table>
<thead>
<tr>
<th></th>
<th>CST (n = 10)</th>
<th>SET (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Basic Laparoscopic Cases$^|$</td>
<td>8 (14)</td>
<td>8 (11)</td>
</tr>
<tr>
<td>No. Intermediate Laparoscopic Cases$^|$</td>
<td>1 (2)</td>
<td>0</td>
</tr>
<tr>
<td>No. Advanced Laparoscopic Cases$^|$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No. Laparoscopic Bariatric Cases$^|$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Participation in a Bariatric Rotation (Yes : No)</td>
<td>1 : 9</td>
<td>0 : 10</td>
</tr>
<tr>
<td>Participation in a MIS Rotation (Yes : No)</td>
<td>1 : 9</td>
<td>0 : 10</td>
</tr>
<tr>
<td>MIS Rotation (months)$^|$</td>
<td>0.2 (0.6)</td>
<td>0</td>
</tr>
</tbody>
</table>

CST – conventional training; SET – simulation-enhanced training; SD – standard deviation; MIS – minimally invasive surgery rotation (participation in intermediate and advanced laparoscopic cases); No. – number. $^\|$ = mean (SD)
Figure 21. Dot plot of the technical skill scores at baseline assessment for the CST and the SET group. Crosses represent means, dots represent individual scores.
Figure 22. Dot plot of the technical skill scores for the CST and the SET group for the 1st Jejunojejunostomy in a live anesthetized porcine model. Crosses represent means, dots represent individual scores.
4.5.2 Secondary Outcome: Knowledge

There was no difference in test scores of knowledge between the SET and CST groups at post-intervention [12.8(2.9) versus 13.9(1.8), p = 0.33). Only 3 of the 10 participants in the SET group read the material provided for the cognitive component of the curriculum.

4.5.3 Secondary Outcome: Technical Skill in the Operating Room

Patient characteristics were similar for SET and CT group: age [42.4(8.5) versus 39.8(10.4), p = 0.58), sex [1 male: 7 female versus 2 male: 8 female, p = 1.0), BMI [46.3(4.3) versus 47.4(8.3), p = 0.73], and ASA [3(0) versus 3.1(0.3), p = 0.39]. There was no significant difference in technical skill of participants in the OR (Table 17, Figure 23). The proportion of operative steps completed without take-over was 0.86(0.16) in the SET group and 0.67(0.26) in the CST group (p < 0.10). Significant improvement in technical skill of participants was observed from the 1st JJ in the live porcine model to the JJ in OR in the SET group [BOSATS: 56.4(11.5) versus 64.1(10.8), p < 0.04, d = 0.69] and the CT group [BOSATS: 46.0(10.6) versus 61.1(8.8), p < 0.01, d = 2.08] (Figure 24 & 25).
Figure 23. Dot plot of the technical skill scores for the CST and the SET group for the jejunojejunosotmy in the operating room. Crosses represent means, dots represent individual scores.
Figure 24. Dot plot of the technical skill scores for the CST group at baseline, on the 1st jejunoojejunostomy in the live anesthetized porcine model and the jejunoojejunostomy in the operating room. Crosses represent means, dots represent individual scores.
Figure 25. Dot plot of the technical skill scores for the SET group at baseline, on the 1st jejunojejunostomy in the live anesthetized porcine model and the jejunojejunostomy in the operating room. Crosses represent means, dots represent individual scores.
4.5.4 Secondary Outcome: Performance in a Simulated Crisis Scenario

The inter-observer agreement was high for each NOTSS skill category and the total element score (Table 19). The SET group scored significantly greater on the post-training assessment in all NOTSS categories, with effect sizes ranging from $d = 1.02$ to $d = 1.36$ (Table 17, Figure 26). Performance of the SET group on a practice scenario was equivalent to the performance of CT group on a post-intervention scenario [Situation Awareness (SA): 8.4(1.9) versus 8.6(2.7), $p = 0.58$; Decision Making (DM): 7.8(2.1) versus 8.2(2.8), $p = 0.64$; Communication and Teamwork (C&T): 8.2(2.0) versus 6.9(1.4), $p = 0.06$; Leadership (L): 8.0(2.2) versus 8.0(2.4), $p = 0.50$; Total Score: 32.4(7.4) versus 31.6(8.7), $p = 0.41$]. The SET group demonstrated a significant within-group improvement in non-technical skill from practice to post-intervention scenario [SA: 8.4(1.9) versus 10.8(1.3), $p < 0.01$, $d = 1.44$; DM: 7.8(2.1) versus 10.5(1.5), $p < 0.01$, $d = 1.54$; C&T: 8.2(2.0) versus 9.2(1.4), $p = 0.09$, $d = 0.55$; L: 8.0(2.2) versus 10.3(1.5), $p < 0.01$, $d = 1.13$; Total Score: 32.4(7.4) versus 40.8(4.2), $p < 0.01$, $d = 1.32$].

Table 19. Inter-observer agreement coefficients for skill categories on the NOTSS scale.

<table>
<thead>
<tr>
<th>NOTSS Skill Category</th>
<th>N</th>
<th>ICC</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation Awareness</td>
<td>37</td>
<td>0.87</td>
<td>0.74 - 0.93</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Decision Making</td>
<td>37</td>
<td>0.86</td>
<td>0.72 - 0.93</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Communication and Teamwork</td>
<td>36</td>
<td>0.90</td>
<td>0.80 - 0.95</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Leadership</td>
<td>36</td>
<td>0.90</td>
<td>0.81 - 0.95</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Total of all element scores</td>
<td>37</td>
<td>0.93</td>
<td>0.87 - 0.97</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

N - number of videos analyzed; ICC – Intraclass correlation coefficient (Two-way mixed-effects model, absolute agreement, average measures); CI – confidence interval, P - level of significance.
Figure 26. Dot plot of non-technical skill scores for post-intervention assessment of the CST and the SET group. Crosses represent means, dots represent individual scores. CST – conventional surgery training; SET – simulation-enhanced training; SA – situation awareness; DM – decision making; CT – communication & teamwork; L – leadership.
4.5.5 Comparison with the PGY 5 Group

Of the 12 final year residents, 9 completed the knowledge test, 9 completed the non-technical skill assessment, and 9 completed the laparoscopic JJ in the OR.

4.5.5.1 Cognitive knowledge

The CST group and the SET group scored significantly less than the PGY 5 group on the knowledge test (Table 17).

4.5.5.2 Technical skill in the operating room

Patient characteristics were similar for the CST, SET and PGY 5 groups: age [39.8(10.4), 42.4(8.5), and 42.3(9.8); p = 0.81], sex [2 male: 8 female, 1 male: 7 female, and 1 male: 7 female; p = 1.00], BMI [47.4(8.3), 46.3(4.3), and 48.8(8.0); p = 0.79] and ASA [3.1(0.3), 3(0), and 3.1(0.4); p = 0.64]. The SET and the PGY 5 group had equivalent technical skill scores (Table 17, Figure 28). The CST group had equivalent technical skill to the PGY 5 group on the BOSATS scale, but scored significantly less on the OSATS GRS scale. The proportion of the operative steps completed without takeover was similar for the CST, SET and PGY 5 group.

Using the predefined minimum level of clinical proficiency (BOSATS ≥ 60%), 2 out of 9 participants in the PGY 5 group were not proficient in the operating room (Figure 27). Technical skills of the PGY 5 group were significantly less than the previously published technical skills of an experienced surgeon group (t(18) = -3.68, p < 0.002) (Figure 28) (Zevin et al., 2013a).

4.5.5.3 Non-technical skill

The SET group scored significantly greater than the PGY 5 group in 3 out of 4 NOTSS categories (Table 17). The CST and the PGY 5 group had equivalent scores.
Figure 27. Dot plot of the technical skill scores for jejunojejunostomy in the operating room for the CST, the SET, and the PGY 5 group. Crosses represent means, dots represent individual scores.
4.6 Discussion

In this study, a comprehensive, simulation-enhanced training (SET) curriculum for a complex laparoscopic operation was developed, and its effectiveness was compared against a conventional surgery residency training (CST) curriculum. The SET curriculum addressed multiple components of surgical competency, including knowledge, and technical and non-technical skills. It was developed in accordance with our prior work using a consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery (Zevin, Levy, et al., 2012), and the educational principles of distributed practice, proficiency-based training and objective assessment. The comprehensive nature of the SET curriculum and its focus on an advanced minimally invasive procedure differentiated this curriculum from other simulation-enhanced training curricula (R. Aggarwal, Ward, et al., 2007; Larsen et al., 2009; Peters et al., 2004). When compared to conventional surgery training, training in the SET curriculum resulted in superior technical performance in a live anesthetized porcine model, superior non-technical skills in a simulated operating room crisis scenario, and equivalent knowledge.

This study had several strengths. First, the prospective, single-blinded randomized design and the use of a conventionally trained group as a control increased the validity and generalizability of the results. The SET group was not compared to a novice “untrained” group; rather, the SET group was compared to the current “status quo” for an intermediate level surgery trainee. Second, inclusion of a component of training on a live anesthetized porcine model allowed for a standardized assessment of technical skill without confounding patient factors. Third, the addition of a PGY 5 comparison group allowed for comparison of a pre-trained PGY 3 or 4 trainee to a PGY 5 resident deemed ready for independent practice. Fourth, the dropout rate of 0-10% for all outcome measures was less than the rates observed in many other randomized trials of simulation-enhanced education (Grantcharov et al., 2004a; Orzech et al., 2012), thereby decreasing the risk of response bias.

4.6.1 Technical Skills Component

Training of a PGY 3 or 4 resident in a SET curriculum was more effective than conventional surgery training. Our results are in agreement with the results of other randomized controlled trials of simulation training in surgery (Palter & Grantcharov, 2012a; Zendejas et al., 2011). Palter and Grantcharov demonstrated that curriculum-trained surgery residents had superior
technical skills in the operating room in comparison to conventionally trained surgery residents (Palter & Grantcharov, 2012a). Zendejas and colleagues reported similar results (Zendejas et al., 2011).

The live anesthetized porcine model was intentionally chosen as a surrogate for a real patient. This model allowed for a safe environment for participants to practice and to make mistakes without the risk of takeover from a supervising surgeon. We set the minimum level of clinical proficiency as BOSATS score ≥ 60%, because with a learning curve of 50-100 cases for a complex laparoscopic bariatric procedure (Zevin, Aggarwal, et al., 2012) we did not expect participants to achieve proficiency of an experienced surgeon (BOSATS > 80%) within the time and financial constraints of this study. Significantly more participants in the SET group were able to demonstrate a minimum level of clinical proficiency in a live porcine model on their 1st and 2nd case suggesting that the SET curriculum resulted in a “pre-trained” mid-level resident. Gallagher and colleagues defined the term “pre-trained novice” as an individual who has been trained using simulation to the point where many of the psychomotor skills and special judgments have been automated (Gallagher et al., 2005). A “pre-trained” resident is expected to have increased cognitive recourses in the operating room to focus on learning the steps of the operation and handling complications rather than using valuable operating room time on the initial refinement of technical skills (Palter, Grantcharov, et al., 2011). Completion of the SET curriculum is expected to standardize the technical skills of surgery trainees to a minimum level of proficiency to maximize the opportunity for safe participation and learning in the operating room.

A large and significant between-group effect size for training in a SET curriculum was demonstrated despite a relatively small time commitment of approximately 2 hours per participant in a simulation laboratory. Larger within-group effect sizes confirmed learning within the SET group. Interestingly, no learning was demonstrated within the CST group despite 4 months of conventional surgery training. Two conclusions can be drawn from these results. First, the exposure to complex laparoscopic surgery in the current training curriculum appears to be fragmented, with only 1 out of 10 participants in the CST group participating in a designated bariatric / minimally invasive rotation during the study period. Participation in the SET curriculum may, therefore, facilitate exposure to and standardize the acquisition of complex laparoscopic skills among mid-level surgery trainees. It may also ensure equal exposure to these skills for all residents irrespective of rotation allocation. Second, the relatively small time
commitment for laboratory training is expected to aid future implementation of the SET curriculum.

The finding of technical equivalence between the SET and CST groups in the operating room was expected, because both groups were trained to a predefined level of proficiency in a live porcine model. The SET group, however, required less time and fewer cases in the porcine model to achieve this goal. If the porcine model can be considered a surrogate for a real patient, this finding suggests that the SET trained participants would require fewer cases in the operating room to achieve proficiency when compared to the CST group. Setting greater cut-offs of proficiency and including a component of overtraining in the laparoscopic box trainer may decrease and/or eliminate the need for live animal training in the SET group (Stefanidis, Scerbo, Montero, Acker, & Smith, 2012).

Comparison of the SET and PGY 5 groups demonstrated equivalent technical performance and equivalent proportion of operative steps completed without takeover in the operating room. This observation suggests that a mid-level surgery resident, without prior experience in complex laparoscopic surgery, can overcome the early part of his learning curve by training in the SET curriculum in the safety of a simulation laboratory. This earlier advancement along the learning curve can maximize his opportunities for participation in complex laparoscopic cases during the senior years of residency training. Our results lend further support to a transition from time-based to a competency-based surgical training (Sachdeva et al., 2011). The finding of inferior operative performance of the PGY 5 group in comparison to the previously published performance of experienced surgeons is consistent with the finding that 23% of graduating residents do not feel fully prepared to practice as an attending surgeon by the end of a 5-year general surgery residency (Coleman, Esposito, Rozycki, & Feliciano, 2013). Our results suggest that introduction of a comprehensive simulation-enhanced curriculum early in the course of residency may decrease the variability in educational outcomes after surgical training and produce more competent, ready-to-practice graduating residents. Implementation of competency-based education will increase the quality of training and will contribute to better utilization of training opportunities and resources, addressing multiple organizational and financial pressures for the current educational system.
4.6.2 Non-Technical Skills Component

Our results show that post-training, non-technical skills of the SET group were superior to the non-technical skills of the CST and PGY 5 groups. Large effect sizes for training in the SET curriculum were observed with a relatively small time commitment. Our results are in agreement with the work of others (Gettman et al., 2009; Knudson et al., 2008). Significant improvements in team performance were demonstrated post training in a quasi-experimental, pre-post study teaching teamwork and communication to senior and junior urology residents (Gettman et al., 2009). In a randomized controlled trial of mid-level surgery residents, participants trained in a simulation-based curriculum scored greater on assessment of crisis management skills and teamwork in real-life trauma resuscitations than residents trained by lecture alone (Knudson et al., 2008).

Similar non-technical skill scores for the CST and the PGY 5 group suggest that experiential learning over the course of residency training does not lead to an improvement in non-technical skills. A growing body of evidence supporting the importance of non-technical skills in the operating room (Christian et al., 2006; Lingard et al., 2004; Mazzocco et al., 2009), and clear recommendations by regulatory bodies (ACGME, 2012) underscore the need for such components in future training curricula. The need for non-technical skills training was also highlighted in a recent survey of General Surgery program directors in Canada and United States (Dedy, Zevin, Bonrath, & Grantcharov, 2013). In this report, only 32% of responding programs offered some form of designated team-training interventions for residents.

4.6.3 Cognitive Component

Our results demonstrated no difference in test scores of knowledge between the SET and the CST groups. This result was expected given that only 30% of participants in the SET group completed their assigned readings. The questions on the test were based primarily on the readings and not on the information presented during the faculty-led seminar. Mandatory completion of reading materials should be introduced when implementing the SET curriculum. Greater scores for the PGY 5 group were expected given the longer duration of exposure to a didactic component of the current surgery training curriculum.
4.6.4 Future Directions

The results of this and other randomized controlled trials (RCTs) (R. Aggarwal, Ward, et al., 2007; Palter & Grantcharov, 2012a; Palter, Orzech, Reznick, & Grantcharov, 2013; Zendejas et al., 2011) confirm that addition of an appropriately developed, SET curriculum to a conventional surgery training program results in improved technical performance in the operating room. We suggest that additional RCTs comparing rigorously developed SET curricula focused on technical skills and conventional surgery training are no longer required nor ethically justified. Future RCTs should focus on head-to-head comparisons of the effectiveness and efficiency of different approaches for performance enhancement in surgery and their impact on educational outcomes. Additional studies are also required to set widely acceptable objective benchmarks for determination of proficiency in complex laparoscopic operations and to compare the cost-effectiveness of this curriculum to conventional surgery training. Multi-institutional studies with prospectively collected patient data will be required to investigate whether training in a SET curriculum leads to improved patient outcomes and safety.
4.7 Limitations

This study had a number of limitations. First, the recruitment rate of 77% could have led to a selection bias; however, this rate is in agreement with rates reported in other randomized controlled trials of simulation-enhanced training (Larsen et al., 2009; Zendejas et al., 2011). Individuals who refused to participate were all PGY 3 and 4 residents in the same program as the study participants and thus were not expected to have a different clinical experience.

Second, this study was conducted at a single, large, surgery residency-training program, and it is unclear if these results can be generalized to training programs with another educational structure and content. Future multi-institutional studies involving all types of programs will be required to demonstrate generalizability of these results.

Third, the decision to allow study participants to progress from a live porcine model to the operating room was made by different preceptors in real-time during the live porcine model training session. It is possible that some preceptors were more stringent, whereas others were less stringent in their assessment. Real-time preceptors scored participants’ technical skills significantly greater than the blinded rater on video review (pg. 138), and this could have been a result of “observer bias”. If a preceptor was familiar with a participant, he could have been influenced by participant’s personality and level of training. Training preceptors to use an objective rating scale and instructing them to use it for declaration of proficiency may lead to greater agreement between the preceptors, as well as between preceptors and an external blinded reviewer. Future studies are warranted to examine the agreement between concurrent live assessment and video-based assessment.

Fourth, the assessment of non-technical skills by an observer who was involved in the training of the participants may have introduced an observer bias. Involving a second observer, blinded to the identity and group allocation of the participants, mitigated this limitation. High inter-observer agreement between the two observers suggested minimal observer bias in the non-blinded rater.

Fifth, the SET curriculum was not designed to address the maturation of surgical judgment. We have shown that completion of the SET curriculum resulted in a “pre-trained” senior resident who may be expected to have increased cognitive resources for higher level learning, such as decision-making and surgical judgment, in the operating room. A component of surgical
judgment (indications for an operation, contraindications, patient selection, intra-operative and post-operative care) was also taught in the cognitive component of the curriculum.

Sixth, PGY 3 and PGY 4 surgery residents were recruited as participants in the CST and the SET groups. There were no differences in the number of basic, intermediate and advanced minimally invasive surgery cases performed by both groups at baseline (Table 16) and the number of operative cases performed during the duration of this study (Table 18). This finding suggested equivalence in operative exposure between the groups. We must acknowledge, however, that the overall clinical skills (knowledge, surgical judgment, decision making, etc.) may be different between a PGY 3 and a PGY 4 surgery resident as there is a whole year of clinical training that separates these individuals. The results of the present study did not allow us to make a comparison between the technical skill of a PGY 3 surgery resident trained in the SET curriculum and a PGY 4 surgery resident trained in the CST curriculum, nor could we compare the technical skill of a PGY 3 and a PGY 4 residents trained in the SET curriculum. Future multi-institutional studies with greater number of participants in the SET and the CST group are required to conduct such subgroup analysis.

Lastly, the design of our study mandated for equivalence of technical skills in the operating room for the CST and the SET group. It can be argued that an ideal study design would take a trainee directly into the operating room without the intermediate step of training in a live anesthetized porcine model. With such a design, we can hypothesize that the difference in technical skills between the SET and the CST groups seen on the first JJ in the live porcine model would also be seen in the real operating room. We believe, however, that in the setting of a research study, it is not ethical to subject a patient to any possible harm by bring an inexperienced participant directly into the operating room without prior training to proficiency in a simulated environment. As a result, we elected to offer an intermediate step of training on a live anesthetized animal as a surrogate for a human patient.
4.8 Conclusions

This prospective, single-blinded randomized controlled trial demonstrated that training in a comprehensive SET curriculum for a complex laparoscopic procedure resulted in superior technical and non-technical skills in comparison to conventional surgery training. Implementation of this SET curriculum into the current surgery training may standardize the training in laparoscopic bariatric surgery, potentially shorten the learning curves in the operating room, and maximize the opportunities for trainees to participate in and learn complex laparoscopic procedures.
Chapter 5

General Discussion
5 General Discussion

The objective of this thesis was to design and validate a comprehensive, SET curriculum for a complex minimally invasive operation. This objective was accomplished using three specific aims. Aim 1 was to develop a framework for design, validation, and implementation of SET curricula in surgery using Delphi methodology and consensus of international experts in surgical and medical education (Chapter 2). Aim 2 was to develop and to demonstrate reliability, validity and feasibility of use of a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery using a hierarchical task analysis, Delphi methodology and international consensus of experienced bariatric surgeons (Chapter 3). Aim 3 was to design a comprehensive, SET curriculum for a complex minimally invasive operation that addressed multiple components of surgical competency (technical skills, non-technical skills, as well as knowledge). A prospective, single-blinded, randomized controlled trial for mid-level surgery trainees was used to compare the effectiveness of this comprehensive curriculum on acquisition of clinical knowledge, technical skills in the operating room and non-technical skills in a simulated operating room (Chapter 4). Our overall hypothesis for the thesis was that curriculum-trained residents will demonstrate superior technical skills, superior non-technical skills, and superior knowledge when compared to conventionally trained residents. The results of our studies support our hypothesis in regard to technical and non-technical skills; however, we did not demonstrate superiority of knowledge in either of the two groups. Discussion sections at the end of each chapter have highlighted the main points of discussion for each study. In this section, I discuss the main findings from each study, and their implications, limitations and relevance to current concepts in surgical education. Directions for future work will be provided in the last subsection of this general discussion.

The consensus-based framework for the design, validation, and implementation of SET curricula in surgery discussed in Chapter 2 served as a theoretical template for the development of a comprehensive, SET curriculum in complex minimally invasive surgery. The first step in this framework was a predevelopment analysis, which consisted of a needs assessment, definition of goals, and objectives, and the desired learning outcomes for a comprehensive SET curriculum. Measures of proficiency were also defined during this step. The needs assessment for our proposed SET curriculum was performed by Palter and colleagues in 2009 (Palter, Orzech, Aggarwal, Okrainec, & Grantcharov, 2010). This needs assessment highlighted the limited exposure of senior surgery residents to advanced minimally invasive procedures. In Palter and
colleagues’ study, only 3 of 14 (21%) senior surgery residents reported participating in laparoscopic foregut and laparoscopic bariatric procedures. As a result, the authors hypothesized that such limited exposure to advanced laparoscopic procedures was not sufficient to develop proficiency in these operations. Indeed, their hypothesis is supported by our finding that proficiency in laparoscopic bariatric surgery is achieved after 50-100 cases (Zevin, Aggarwal, et al., 2012). Therefore, by completing of a comprehensive SET curriculum for laparoscopic bariatric surgery, surgery trainees will have exposure to and an opportunity for deliberate practice of technical skills required in complex laparoscopic surgery. Practice in a simulation laboratory can shorten the learning curves in the operating room (Larsen et al., 2009), thereby allowing a mid-level surgery trainee to participate actively in the technical aspects of advanced laparoscopic cases. Once the predevelopment analysis was completed, we turned out attention to the development of such a comprehensive SET curriculum for laparoscopic bariatric surgery.

Proficiency-based training is one of the cornerstones of simulation-based training of technical skills. Such training requires an objective assessment scale to evaluate the technical skills in the operating room and in the simulation laboratory. Such a scale did not exist for laparoscopic bariatric surgery. Consequently, the second aim of this thesis and the next step in the curriculum framework was to develop and demonstrate feasibility of use, as well as the reliability and validity of a scale for objective assessment of operative performance in LRYGB. This study was described in Chapter 3. The first step in the development of the scale was to deconstruct the LRYGB into its component steps using a hierarchical task analysis. The outcome of that hierarchical task analysis was an extensive list of operative steps required to perform a complete LRYGB. Delphi methodology was then used to survey an international panel of experts in laparoscopic bariatric surgery in order to identify the operative steps that were relevant to training and assessment of technical skill in LRYGB. The resultant list of steps (Table 11) can be viewed as a “training itinerary” for LRYGB. This itinerary can be a useful resource for surgical educators and companies interested in developing simulator technology. It may guide the development of novel training strategies and simulation models for laparoscopic bariatric surgery. For example, the itinerary highlighted the creation of handsewn, linear stapled and circular stapled gastrojejunostomy as an important component for training and assessment of technical skill in LRYGB. At the present time, there are no synthetic or cadaveric models to teach and assess these skills in a simulation laboratory. Future research should focus on the development and validation of such models. Furthermore, this itinerary may also be used to develop strategies for mental rehearsal and mental practice for maintenance of technical skills;
such strategies defined as “the cognitive rehearsal of a task before performance” have been shown to enhance technical skills and decrease stress in the operating room (Arora, Aggarwal, et al., 2011).

The list of operative steps generated from the hierarchical task analysis was also used to develop the Bariatric Objective Structured Assessment of Technical Skill (BOSATS) scale by assigning anchoring descriptors of performance to each operative step. One of the strengths of the BOSATS scale is its generalizability to most operative approaches to the LRYGB. This generalizability is expected to increase the acceptance and use of this scale in surgery residency and bariatric fellowship training programs. The other strengths of the BOSATS scale include its inter-rater and test-retest reliability, both of which are greater than those reported for OSATS (J. A. Martin et al., 1997) and GOALS (Vassiliou et al., 2005) scales; and the high construct and concurrent validity. Most importantly, the BOSATS scale was also shown to be feasible to use, requiring approximately 30 minutes to evaluate a 120-minute video-recorded procedure. All of these psychometric properties of the BOSATS scale are expected to enhance its acceptance and use for the objective assessment of operative skills in LRYGB, proficiency-based training and deliberate practice. In its current form, the BOSATS scale is not suitable for use in summative, high-stakes assessment. The scores on this scale should be used for constructive feedback to surgeons rather than in high-stakes certification and re-certification decisions. Additional work is required to define cut-off values with high sensitivity and specificity for use in the more high-stakes decisions concerning certification and re-certification. Cusimano provided an excellent summary of available methodologies for setting cut-off values for high stakes examinations (Cusimano, 1996). One approach is to use the borderline-group method, which centers on the assessment of performance of the examinee. In this method, a group of judges such as experienced bariatric surgeons identify a group of participants who they consider to have a borderline performance. The mean BOSATS score for that group would be considered as a cut-off score for a minimum acceptable level of technical proficiency in the LRYGB.

With the predevelopment analysis completed and the BOSATS scale developed, the next aim in the thesis and the next step in the theoretical framework was to develop the technical, non-technical and cognitive components of the SET curriculum for laparoscopic bariatric surgery. As I discussed previously in the introduction chapter, acquisition of cognitive knowledge should be the first step in technical skill training. Indeed, Fitts and Posner suggested that “cognition” is the first stage of acquisition of technical skill (Fitts & Posner, 1967). At this stage the trainee begins
to gain an insight into the performance of a technical task through instructor explanation and demonstration. Van Herzeele and colleagues provided experimental evidence in support of this theory (Van Herzeele et al., 2008). In their study, 20 junior vascular surgeons were sequentially allocated to Group 1 (a 45-minute, one-to-one didactic training session covering indications for treatment, relevant anatomy, potential errors, dangers of fluoroscopy, etc.) or Group 2 (no cognitive training) and were asked to perform a virtual reality iliac angioplasty and stent procedure. Group 1 used a more appropriate stent size and had less residual stenosis in the artery compared to Group 2. The authors concluded that cognitive training lead to a significant improvement in endovascular technical skills. The cognitive component of the SET curriculum in laparoscopic bariatric surgery was designed to provide a trainee with the relevant knowledge of pre-operative, intraoperative and post-operative aspects of bariatric patient care. This cognitive component was administered via a 2-hour interactive, faculty-led seminar, combined with self-directed readings. This approach to the delivery of the cognitive component has been used effectively in other studies of SET curricula (Palter & Grantcharov, 2012a; Zendejas et al., 2011).

The technical component of our SET curriculum was structured around training to proficiency in technical skills using a laparoscopic jejunojejunostomy model. Proficiency level was set as the mean score of experienced bariatric surgeons for the creation of a laparoscopic jejunojejunostomy as part of the LRYGB (Zevin et al., 2013a). The technical component focused primarily on the creation of jejunojejunostomy, because surgery trainees most frequently perform this component of the LRYGB; minimally invasive or bariatric surgery fellows usually create the gastric pouch and the gastrojejunostomy. However, the technical skills required to create a jejunojejunostomy are the same skills required to perform a laparoscopic small bowel resection for a tumor or to perform a laparoscopic right hemicolectomy. We hypothesize that skills acquired during the creation of laparoscopic jejunojejunostomy will be transferable to other minimally invasive operations; however, this hypothesis has to be confirmed in future studies.

The non-technical component of the SET curriculum addressed such skills as teamwork, communication, decision making, situation awareness and leadership. These skills were taught via a 1-hour, expert-led, interactive seminar and a 15-minute, simulated intra-operative crisis scenario with a 15-minute debriefing session. This approach to non-technical skills training was based on the best-available evidence for non-technical skills training in a surgery residency (Dedy, Bonrath, et al., 2013). This comprehensive nature of our SET curriculum (cognitive,
technical and non-technical skill component) is its main strength. To our knowledge, this curriculum is the first of its kind to comprehensively address all components of surgical competency for a complex minimally invasive operation. Several other simulation-based training curricula have combined cognitive and technical skills training (Palter & Grantcharov, 2012a; Swanstrom et al., 2006); however, none of these studies have addressed all components of surgical competency.

The last aim of this thesis was to compare the effectiveness of our SET curriculum to conventional surgical training. This aim corresponds to the curriculum validation, evaluation, and improvement step of the curriculum framework. Indeed, as described in chapter 4, we were able to confirm our initial hypothesis by showing that SET curriculum-trained residents possessed superior technical skills and superior non-technical skills when compared to conventionally trained residents. Similar results were reported by other groups for ex-vivo training in technical skills in basic and intermediate laparoscopic operations (R. Aggarwal, Ward, et al., 2007; Palter & Grantcharov, 2012a; Zendejas et al., 2011), as well as non-technical skills (Dedy, Bonrath, et al., 2013). We were not able to demonstrate a significant difference in knowledge between SET curriculum-trained and conventionally trained groups, which differed with the results of other groups (Palter & Grantcharov, 2012a; Palter et al., 2013). In a single-blinded, randomized controlled trial of an ex vivo training curriculum for laparoscopic colorectal surgery, Palter and Grantcharov were able to show a significant improvement in knowledge on a multiple-choice test for the curriculum trained group in comparison to the conventional group (p < 0.05). The multiple-choice knowledge test used in that study was based primarily on the information presented during the faculty-led seminar session; whereas in our study, the knowledge test was based primarily on the material covered in the self-directed readings and not on the information presented during the faculty-led seminar session. Our finding of no difference on a knowledge test between the SET curriculum and conventionally trained group was not expected. However, when one considers the fact that only 30 percent of curriculum-trained group completing the self-directed readings, the finding of no difference between the groups can be explained. Moreover, the finding of our study leads us to advocate for mandatory completion of the cognitive component of the curriculum in the implementation phase.

The last step in the theoretical framework for curriculum development included the evaluation and improvement of the curriculum. Both of these tasks should be carried out after the implementation of the SET curriculum into a general surgery residency-training program. Data
on the perceptions, improvement and retention of knowledge, technical and non-technical skills of trainees should be collected prospectively at regular time intervals. With these data for regular review, program administrators will be able to evaluate and improve this curriculum repeatedly.
5.1 Implementation of a Comprehensive Simulation-Enhanced Training Curriculum

In this thesis we have demonstrated that training in a comprehensive SET curriculum in LRYGB resulted in superior technical and non-technical skills when compared to conventional surgery training. The next step is to translate our knowledge from the laboratory to the clinical practice by implementing this SET curriculum into the surgery training program in Toronto and potentially into other surgery training programs outside Toronto.

Several factors must be considered to ensure successful implementation of the SET curriculum. First, the structure and the comprehensive nature of the SET curriculum must be maintained during the implementation phase. It would not be possible to predict whether the findings reported in this thesis would be replicated during the implementation of the SET curriculum if one of the components of the curriculum were to be removed and/or substantially modified. We do, however, acknowledge that all curricula should be dynamic and responsive to new ideas, technologies, and knowledge.

Second, training within the SET curriculum must remain proficiency-based rather than time-based. From a logistical perspective, it is also important to estimate the maximum realistic amount of time that may be required by surgery trainees to reach proficiency in a laboratory setting, as well as the maximum realistic time commitment for faculty participation in the SET curriculum. The trainee’s time commitment will depend on his level of technical skills at baseline and on his ability to acquire complex minimally invasive surgery skills. Data from this thesis on the maximum amount of time required to achieve a predefined level of technical skill proficiency can be used as a guide by program administrators to estimate of the recourses that may be required to run the SET curriculum.

Grantcharov and Funch-Jensen described four distinct patterns of learning curves for basic, minimally invasive surgical tasks (Grantcharov & Funch-Jensen, 2009). In their study of 37 novice surgery residents performing basic laparoscopic surgical tasks on a VR surgery simulator, approximately 6 percent of participants were proficient from the beginning of training, 70 percent were able to achieve proficiency in 2-9 repetitions on a simulator, 16 percent showed improvement, but were unable to achieve proficiency in 10 repetitions; and 8 percent showed no tendency for improvement. Consequently, within the SET curriculum we would expect some trainees to fall into the “proficient from the beginning of training” group, and others to fall into
the “no tendency for improvement” group. In our study, 10 mid-level surgery residents were allocated to the SET curriculum group. All of these residents achieved a predefined level of proficiency on the laparoscopic jejunojejunostomy box trainer model. The mean duration of time to reach proficiency in a box trainer was 2.1 (0.8) hours. The fastest time was 0.9 hours and the slowest time was 3.2 hours. There were no residents who failed to improve despite repeated attempts. This finding may be a reflection of a small group size within the SET group (one would expect 0.8 participants to show failure to improve within a group of 10 participants), or a reflection of the individualized one-on-one feedback that was provided to each resident within the SET group. The results of the study by Grantcharov and Funch-Jensen can be confirmed during the implementation of the SET curriculum, as greater number of trainees will be expected to participate.

The mean duration of time spent practicing on a live anesthetized animal model was 1.7 (0.7) hours and the greatest duration of total practice time was 2.5 hours. Adding the greatest duration of practice time on the laparoscopic jejunojejunostomy in the box trainer to the greatest duration of practice time in the live anesthetized porcine model resulted in a total time commitment of 5.7 hours for the technical skills component of the SET curriculum. The total time commitment for the non-technical component was 1.5 hours (0.5 hours for the simulated intraoperative crisis scenario with a debriefing session and 1.0 hours for the interactive, faculty-led seminar on nontechnical skills in the operating room). The total time commitment for the cognitive component of the SET curriculum was 2.0 hours for the faculty-led seminar and additional time for the self-directed reading component.

Adding up the maximum time requirements to complete each of the components of the SET curriculum resulted in a total time commitment of 9.2 hours. Therefore, a program administrator who is planning to implement the SET curriculum should allocate a maximum of 9.2 hours of educational time for each resident to complete the SET curriculum. By investing 9.2 hours of educational time in the SET curriculum we can expect for an intermediate level surgery trainee (e.g. PGY 3) to gain equivalent technical skills and superior non-technical skills to a current graduating PGY 5 resident subjected to our current conventional educational scheme that does not utilize the SET curriculum. In order to adhere to a schedule of distributed practice, the comprehensive SET curriculum might best be administered over the course of 5 weeks with 2 hours of training per week.
The time commitment of 9.2 hours for our SET curriculum is similar to the time commitment reported in other studies of proficiency-based simulation training (R. Aggarwal, Ward, et al., 2007; Ahlberg et al., 2007a; Palter & Grantcharov, 2012a; Scott et al., 2008). Palter and Grantcharov reported an average time of 8.3 hours per trainee to complete their curriculum, which contained a cognitive and technical skill component (Palter & Grantcharov, 2012a). Scott and colleagues reported an average of 9.7 hours to achieve proficiency on the technical skill component of the Fundamentals of Laparoscopic Surgery curriculum (Scott et al., 2008). The two randomized controlled trials that assessed the effect of proficiency-based training on a VR simulator on the operating room performance, reported that the time spent training on the simulator ranged from 1.1 hours to 7.0 hours to achieve expert levels of proficiency (R. Aggarwal, Ward, et al., 2007; Ahlberg et al., 2007b). Based on these data, the time commitment required to complete our comprehensive SET curriculum was well within the range of what is currently described in the literature.
5.2 Cost of Training in a SET Curriculum

The total cost of training 10 participants in a comprehensive SET training curriculum was about $23,800 CAD ($5,700 CAD for training in a box trainer, $16,100 CAD for training in a live anesthetized porcine model, $1,000 CAD for non-technical skills training, and $1,000 CAD for cognitive training). This cost includes rental of space in the simulation center, purchase or rental of surgical equipment, and the times required of the faculty, residents and technical personnel. The cost of performing one laparoscopic jejunojejunostomy in the operating room by all 10 curriculum-trained residents was about $23,200 CAD. Therefore, pre-training of the 10 mid-level general surgery residents to a technical level of a PGY 5 resident in a simulation laboratory was $600 CAD more expensive than having each resident perform one anastomosis in the operating room. From our results in Chapter 4, PGY 5 residents performed an average of 16 advanced laparoscopic cases and 42 laparoscopic bariatric cases by the end of their training. The approximate cost of performing 42 laparoscopic jejunojejunostomies in the operating room by one resident would be $97,300 CAD. Training within the SET curriculum was expensive, but training in the operating room would be even more expensive. For the purpose of the above calculation we used the cost of training for a SET resident who has already overcome the initial part of the learning curve in a simulation laboratory. A PGY 3 resident who has not been previously exposed to SET training would be expected to be slower and less proficient when compared to a SET resident. As a result, the cost of training a simulation-naïve PGY 3 to a level of a graduating PGY 5 in the operating room is expected to be greater than $97,300 CAD.

Ultimately, a cost-effectiveness study comparing a SET-trained group and a simulation-naïve conventional surgery-training group is required. With our evidence of the cost-effectiveness of this SET approach, surgical educators will have the necessary leverage to persuade university departments and hospitals to invest into the comprehensive SET curriculum presented in this thesis.
5.3 Practicalities of Simulation-Enhanced Training in the Real World

There are several barriers to widespread incorporation of simulation-enhanced training in surgical education. Limited availability of surgery faculty to teach outside of the operating room is one of these barriers. The unfortunate reality in North America is that surgery faculty often get paid only when they are directly involved in patient care. Understandably, surgeons are reluctant to trade off their time in the operating room for teaching in a SET curriculum.

There are several plausible solutions to this barrier. Transitioning from a “fee-for-service” to a salary-based system of reimbursement may permit surgery faculty with an academic appointment to spend more time away from the operating room instructing, supervising and evaluating trainees in a simulation laboratory. Alternatively, teaching effectiveness scores and demonstration of commitment to resident education can be used as criteria for academic promotion. For example, surgery faculty who spend a substantial amount of time teaching outside of the operating room may be promoted from an assistant to an associate professor based on their teaching effectiveness and commitment to resident education.

Another solution may be to use non-physician personnel as instructors in a simulation laboratory. A randomized controlled trial of 49 medical students that compared laboratory teaching of basic technical skills by a non-physician skills coach and a faculty surgeon (Kim et al., 2010) showed no difference in technical skill between both training groups on post-intervention testing. The use of non-physician coaches may ease the burden on the surgery faculty, while providing similar quality of instruction for trainees.

Another barrier to the widespread incorporation of simulation-enhanced training in surgical education is the difficulty in removing trainees away from the manpower demands of patient care. Physician extenders, physician assistants and nurse practitioners can be effectively used to enhance patient care and relieve surgery residents from routine and repetitive patient care tasks that may have limited educational benefit. The University of Toronto Orthopedic Surgery Program is a great example of how physician extenders can be used to decrease the patient care workload for surgery trainees, which in turn gives trainees the opportunity to participate in simulation-enhanced training (Alman, Ferguson, Kraemer, Nousiainen, & Reznick, 2013).

Another barrier to widespread incorporation of simulation-enhanced training in surgical education is the opportunity cost of taking surgery residents out of the clinical environment to
participate in simulation-enhanced training. A surgery trainee may potentially miss out on relevant clinical experience by participating in a SET curriculum. This opportunity cost must be acknowledged and addressed by maximizing the educational opportunities for surgery trainees during their training. Current surgery residents spend a substantial amount of time on tasks of limited educational value (e.g., completion of home-care forms, and dictating numerous discharge summaries). The use of physician extenders to complete these low educational value tasks may permit surgery trainees to capitalize on educational opportunities in both the clinical and simulation setting.
Chapter 6

Limitations
6 Limitations

Detailed limitations of each study described in this thesis were included in the chapter describing each study. In this section, I will highlight the most important limitations. The two main limitations of the study regarding the consensus-based framework for design, valuation and implementation of SET curricula in surgery (Chapter 2) were the non-uniform geographical distribution of the participants in the Delphi panel and the recruitment of members of the ASSET (Alliance for Surgical Simulation Education and Training) group as the participants in the Delphi panel. The majority of participants in the Delphi questionnaire were from North America, whereas continental Europe, Asia, and South America were not well represented. Consequently, our results likely reflect an Anglo-American point of view and should be interpreted in the context of an Anglo-American education format. Selecting members of the Delphi panel from the ASSET group may have also biased our results; however, provided that there is no exact criterion in the literature concerning the selection of Delphi panel members (Hsu & Sandford, 2007) we felt that selecting individuals who are interested and passionate about surgical education may increase the response rate to the online questionnaire. It might be beneficial to conduct a similar Delphi survey in continental Europe, Asia and South America, and to pool the results from each continent to have a more generalizable framework. Despite this limitation, however, the current framework is generalizable to most surgical specialties and surgical procedures performed in countries with an Anglo-American education format.

The main two limitations of the study on the development, feasibility, validity, and reliability of a scale for objective assessment of operative performance in LRYGB (Chapter 3) were the recruitment of members of the ALTAS (Advanced Training in Laparoscopic Abdominal Surgery) group as the participants in the Delphi panel and our inability to demonstrate construct validity for each of the components of the Bariatric Objective Structured Assessment of Technical Skills (BOSATS) scale. Selecting members of the ATLAS group as members of the Delphi panel and recruiting greater number of participants for round 2 versus round 1 of the Delphi questionnaire may have introduced bias into our results. These strategies for Delphi panel selection were applied in an effort to maximize the response rate to the online questionnaire and to achieve consensus on the steps of the LRYGB for inclusion into the BOSATS scale. Indeed, the response rate for the study in Chapter 3 was greater than the response rates reported for similar studies in the literature (pg. 129). Furthermore, selecting the Delphi panel members from the members of the ATLAS group increased the feasibility of
conducting the study in Chapter 3 by ensuring that the panel was knowledgeable in laparoscopic bariatric surgery and was interested in developing the BOSATS scale.

Another limitation of the study in Chapter 3 was a result of our inability to evaluate each of the components of the BOSATS scale using a video-recording of an intra-abdominal view of the LRYGB operation. For example, it was not possible to evaluate the components of the BOSATS scale that addressed patient positioning and padding of pressure points, as these were not recorded on the intra-abdominal view of the operation. Future research studies that will evaluate the reliability, validity and feasibility of use of the BOSATS scale in real time rather than on retrospective video review should address this limitation. Furthermore, such studies may also answer the question of whether BOSATS scores obtained in real time are in agreement with BOSATS scores obtained on retrospective review. If the real-time scores and the scores from retrospective review do not agree, generalizability theory may be used to identify the source of measurement error by calculating the proportion of variance in the BOSATS score that is attributed to the participant, the rater or the items on the scale. We hypothesize that real-time assessment can introduce observer bias, because the assessor may know the identity and training level of the resident. In Chapter 4, we have provided some preliminary results in support of this hypothesis (BOSATS scores for laparoscopic jejunojejunostomy assessed in real-time were significantly greater than those assessed on retrospective and blinded review).

The BOSATS scale in its current form was designed for use in formative rather than summative assessment. It can be used for identification of specific areas of weakness, provision of specific targeted feedback, and facilitation of deliberate practice. Additional studies are required to set specific cut-off values of proficiency for use in high-stakes assessments, such as certification and recertification. The methodology for such research studies can be modeled on the study by Fraser et al. that described setting passing scores for the technical component of the Fundamentals of Laparoscopic Surgery curriculum (Fraser et al., 2003).

The main limitation of our prospective, single-blinded randomized controlled trial of a comprehensive SET curriculum (Chapter 4) was the conduct of the study in a single, large surgery training program. As a result, it is not clear whether our results can be generalized to other surgery training programs with different educational structure and content. This limitation can be addressed in future multi-institutional studies involving programs with different educational structures and contexts. Multi-institutional studies will be required to answer the
question of whether patient outcomes (intra-operative and post-operative complications) are improved by training with our comprehensive SET curriculum in laparoscopic bariatric surgery.

It is also important to acknowledge that the results of our study in Chapter 4 do not suggest that surgery residents who complete the SET curriculum can operate independently on patients. Our results suggest that completion of the SET curriculum resulted in a “pre-trained” mid-level surgery resident that possessed the necessary technical and non-technical skills to maximize his learning in the real operating room. Currently, a surgery resident must complete 5 years of surgery residency training, demonstrate satisfactory achievement of all CanMEDS competencies, and pass the Royal College of Physicians and Surgeons of Canada Licensing Examination in order to be granted privileges for independent practice.

Lastly, it is also important to acknowledge that the wealth of knowledge related to non-technical skills training and crisis management from the fields of anesthesia, emergency medicine, and aviation was not included in this thesis for pragmatic considerations. The objective of this thesis was to design and validate a comprehensive, simulation-enhanced training curriculum for a complex minimally invasive operation, therefore extensive review of the literature on non-technical skills and crisis management from non-surgical disciplines was beyond the score of this thesis.
Chapter 7

Future Directions
7 Future Directions

A number of potential avenues for future work have already been suggested at the end of each chapter in this thesis. Herein, I will highlight what I consider to be the most interesting ones. In regard to the consensus-based framework for design, validation, and implementation of SET curricula in surgery, future studies should aim to define the appropriate intervals for retraining of cognitive, technical and non-technical skills. Such intervals can be defined following the implementation of a SET curriculum like that described in Chapter 4. Testing sessions for knowledge, technical and non-technical skills can be set up at regular intervals, such as 1, 3, 6, 9 and 12 months post completion of curriculum training. In this manner, researchers will be able to document the onset and rate of degradation of knowledge, and technical and non-technical skills. The appropriate time for re-testing of skills acquired in a SET curriculum is yet to be defined in the literature. Bonrath and colleagues reported that medical students retain learned technical skills for at least 6 weeks after completion of training (Bonrath et al., 2012). Stefanidis and colleagues reported that minimally invasive technical skills might be retained for up to 5 months post completion of simulation training (Stefanidis et al., 2005). After determining the timing of degradation of skills, it will be possible to develop a “refresher” curriculum for maintenance of training. Residents who demonstrate a substantial loss of knowledge, and technical and non-technical skills at the time of testing may be asked to complete a “refresher” curriculum prior to returning to the operating room.

Future work on the BOSATS scale described in Chapter 3 may follow three different paths. The first path for future work on the BOSATS scale would be to confirm the feasibility of use, validity and reliability of the BOSATS scale during real-time intraoperative assessment of technical skill during LRYGB. It will be interesting to determine if the scores for real-time intraoperative assessment are in agreement with the scores obtained on retrospective video review by a trained rater. Real-time intraoperative ratings will give us an opportunity to demonstrate construct and concurrent validity for items on the BOSATS scale which require more global evaluations, such as the external view of the patient (patient positioning, placement of trocars, etc.).

The second path for future work on the BOSATS scale would be to investigate if rater training is required prior to the use of the BOSATS scale. In the era of finite resources, an objective
assessment scale that can be used by a non-trained, non-clinician rater would markedly enhance the feasibility of use and transferability of the BOSATS scale.

The third path for future work on the BOSATS scale may center on setting sensitive and specific cut-off scores for proficiency for use in high-stakes assessments, such as certification and re-certification. Several methodological approaches are available for standard setting, including the “borderline group” method described by Cusimano (Cusimano, 1996), and the methodology for setting cut-off scores for the MISTELS system described by Fraser and colleagues (Fraser et al., 2003).

Several future studies can build on the results of the prospective, single-blinded randomized controlled trial described in Chapter 4. One study may focus on the implementation and evaluation of the comprehensive SET curriculum in a general surgery training program. The findings from such a study will help us to identify potential barriers to curriculum implementation, such as cost, lack of equipment and faculty resources, and lack of resident participation. Once the SET curriculum has been successfully implemented in one surgery training program, future studies may investigate the implementation of this curriculum in other surgery training programs with different educational structure and content. Each implementation should be accompanied by prospective collection of data on the educational and patient outcomes. Such outcomes data may permit us to answer the questions of whether the SET curriculum is generalizable and transferable to other surgery training programs, and whether simulation-enhanced training results in improved patient outcomes. Such future work will require full support of department chairs, program directors, hospital administration, database managers, and surgical educators.

Future work may also address the topic of cost-effectiveness of the SET curriculum training. I have already touched on this topic in the aforementioned section; however, herein I will provide some methodological detail. A cost-effectiveness study may be set up to compare two groups of surgery residents: a SET curriculum trained group (group 1), and a simulation-naïve group who will be trained entirely in the operating room (group 3). Each group will be required to achieve the same level of proficiency in technical skill as measured on the BOSATS scale. The results of such a study would permit us to compare the cost of training in a comprehensive SET curriculum versus the cost of training in a conventional surgery training program.
Lastly, based on the results reported in this thesis, I would suggest that future randomized controlled trials comparing rigorously developed SET curricula for technical skills training and conventional surgery training are no longer required, nor ethically justified. Future research should instead focus on the head-to-head comparisons of the effectiveness and efficiency of different approaches for performance enhancement in surgery, as well as their impact on educational outcomes.
Bibliography


