Do Fundamentals of Laparoscopic Surgery (FLS) and LapVR Evaluation Metrics Predict Intra-Operative Performance?

by

Sarah Steigerwald

A Thesis Submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Surgery

University of Manitoba

Winnipeg

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to extend my gratitude and thanks to my supervisor Dr. Ashley Vergis. He offered invaluable support and mentorship, and without his patient guidance and encouragement this research project would not have been possible.

I would also like to thank Dr. Jason Park for his added insight, ideas and assistance with this project.

Furthermore, I extend my appreciation to all of the residents that took time out of their busy schedules to participate in this study.

Finally, I would like to thank my family and friends for their endless patience and support throughout this entire project.

Abstract

Background: Considerable resources have been invested in low and high fidelity simulators in surgical training. The purpose of this study was to establish construct and predictive validity for the Fundamentals of Laparoscopic Surgery (FLS, low fidelity box trainer) and LapVR (high fidelity virtually reality) training systems and to establish construct validity for two operative ratings scales (GOALS and OSATS) in our center.

Methods: 26 participants performed tasks from the FLS program and the LapVR simulator as well as a human laparoscopic cholecystectomy. Performance was evaluated using FLS and LapVR metrics, and the previously validated global rating scales.

Results: GOALS and OSATS demonstrated construct validity. Construct and predictive validity was also strongly demonstrated for FLS tasks but only indeterminately for LapVR.

Conclusions: The high cost LapVR remains experimental in resource constrained training programs. Efforts should be focused on utilizing the well-validated, lower cost FLS video trainer for assessment of laparoscopic skills.

Chapter 1 Introduction

The traditional Halstedian (1) approach to surgical teaching of, “see one, do one, teach one”, has fallen out of favour over the last two decades. Laparoscopic surgery, reduced resident work hours, increased costs of operating room time, increased public awareness and the ethics of learning basic skills on patients have led to a re-evaluation of surgical education and how surgical skills are taught and assessed in trainees. This has fueled investigation into the use of basic skills laboratories and surgical simulation as methods of training and assessing residents.

The aviation industry has a long history of utilizing simulation for both training and assessment purposes (2, 3). Krummel (4) defined simulation as “a device or exercise that enables the participant to reproduce or represent, under test conditions, phenomena that are likely to occur in actual performance”. Simulation represents a simplified reality that need not include every possible detail or reproduce anatomy with high fidelity to be effective (5-7). Simulated training and assessment environments allow for practice in a realistic setting without the inherent risk of harming others or oneself. This is particularly important when the target performance involves high stake potentially life-threatening situations, such as with surgery.

A. Laparoscopy

With the introduction of laparoscopy and minimal access surgery (MAS) in the 1980s, patients experienced the benefits of smaller incisions, shorter hospital stays and decreased postoperative pain. However, it quickly became apparent that laparoscopic surgery was not without risk and potential serious complications. The skills required to perform laparoscopic surgery are unique and different than those of traditional open surgery. Laparoscopic surgery uses optical systems that provide monocular vision. This eliminates depth perception, and surgeons must depend on other cues such as light and shading to recreate a ‘stereoscopic’ environment (8). Additionally, the video image is magnified and projected onto a monitor that is not aligned precisely with the surgical target.

Laparoscopy uses long instruments and trocars; these amplify movements and tremor, and are more difficult to control than traditional instruments. As the trocars are fixed to the anterior abdominal wall, there is a decreased range of motion of the instruments. Passing instruments through trocars also results in the fulcrum effect, whereby the surgeon’s hand must be moved 180° opposite the direction of the desired movement of the tip of the instrument (9). The added length of the laparoscopic instruments also significantly dampens tactile sensation. Surgeons rely heavily on their sense of touch to define tissue planes, pathology and the resistance required to secure knots. Cues such as touch and the interaction of specific instruments with tissue must be learned to help maintain the necessary tactile feedback (9).

The potential for devastating complications combined with the unique skill set required to perform laparoscopic surgery prompted the surgical community to reconsider the approach to training and assessment of these skills (10). The need for the demonstration of technical competence by both surgical trainees and practicing surgeons was highlighted. Societies and regulatory bodies such as the Society of Gastrointestinal and Endoscopic Surgeons (SAGES) and the European Association of Endoscopic Surgeons (EAES) stipulated minimum requirements for those performing laparoscopic training with an emphasis on training in both the operating theatre and in skills courses (10-12).

B. Work hour restrictions

In 2003, the Accreditation Council for Graduate Medical Education (ACGME), the professional body responsible for accrediting residency-training programs in the United States, implemented an 80-hour resident workweek restriction (13). In Canada, while there are currently no pan-Canadian limitations on resident duty hours, there are restrictions within each province. Manitoba has a weekly duty hour limit of 89 hours per week. In Quebec, residents' duty hours are limited to 72 hours per week. As of June 2011, they have a limit of 16 consecutive hours per day (14). This may be implemented in other jurisdictions in the next few years. These reduced hours combined with the challenging skill set afforded to laparoscopic surgery, make training of laparoscopic skills and demonstration of laparoscopic competency outside of the operating room even more important.

C. Laboratory simulation based training and assessment

Laboratory simulation based skills training and assessment for laparoscopic surgical skills include lower fidelity physical box/video trainers and higher fidelity virtual reality trainers. The physical box trainers use actual laparoscopic instruments and an optical system that allows the trainee to perform tasks under videoscopic guidance (3). Virtual reality trainers use computer-generated instruments to perform computer-generated tasks. There is a spectrum of machines available now, with some of the newer ones incorporating haptic feedback.

Over the last two decades, much research has been performed studying the validity and reliability of simulation based training for laparoscopic skills. Training in lab-based settings results in improvement of skills both in the simulation setting and the operating room. Simulation is now established as an effective method of skill acquisition (15-33).

Video trainers have been incorporated into the assessment of laparoscopic skills via the Fundamentals of Laparoscopic Surgery (FLS) program. FLS is a simulation based assessment and certification program developed by SAGES (34). It has been thoroughly validated (8, 15, 16, 34-43) and is a requirement to be eligible for certification by the American Board of Surgery (34, 44).

Virtual simulators have also been shown to be capable of assessing the psychomotor skills necessary for laparoscopic surgery (25, 45-51), although there is less research in this area than with video trainers.

To our knowledge no investigation has compared physical box trainers to virtual reality systems head to head for operative assessment purposes at the initiation of this study. This is important as most residency training programs and surgical centers have limited funds and resources. If one form of simulation is superior to the other at assessing laparoscopic skill, resources could be directed accordingly. Additionally, with increasing work hour restrictions there is less time available to train residents to the same end product. This time needs to be used as efficiently as possible.

Before looking further into the value of simulation for assessment of intraoperative laparoscopic skills, we must establish the framework that makes an effective assessment tool.

Chapter 2 Assessment of Surgical Skills

Before any assessment tool is implemented into widespread use, there must be a method to evaluate its product. Generations of medical educators have outlined considerations that guide decisions about developing the most appropriate method for assessing a learned skill (52-54). Considerations include: Why is the assessment being performed? What should be assessed? What are the essential features of an ideal assessment tool? How will the assessment be categorized and judged? Who should perform the assessment? When should the assessment occur? (52)

A. Why is the assessment being performed?

It is of the utmost importance to establish the purpose of the assessment. The acquisition and refinement of technical skills are central to surgical teaching (55). Reliable methods of evaluating operative performance are essential for surgical training programs. However, the teaching and assessment of technical ability and competence has lagged behind other educational endeavors in surgical training programs (3). With the unique skill set afforded to laparoscopic surgery, it is especially important to ensure that surgical trainees are adequately trained in this area. Assessments provide feedback to the learner on their strengths and weaknesses, identifying areas requiring remediation or advancement potential. Assessment can also establish if a trainee is ready to be licensed for independent practice, and help to maintain academic standards by evaluating ongoing competency. Establishing the purpose of the assessment allows the educator to delineate the framework with which to define the assessment method.

B. What should be assessed?

By establishing the purpose of the assessment, we can then look at what should be assessed. Ideally, with the purpose of the course or program and assessment in mind, educational objectives would have been set before the course or program was implemented. The assessment tool should reflect these objectives as they serve to facilitate the same educational product (53). Educational objectives can be broadly classified into three domains: knowledge, skills and attitudes. Knowledge objectives are those that address cognitive measures. These encompass a wide range of items such as being able to recall specific facts, to integrating processes for problem solving. Attitude objectives reflect humanistic qualities and relate to an individual's approach to medicine, patients, peers and the allied health care team. Skills objectives involve the psychomotor and technical skills that are required to be a competent clinician (54).

C. Requirements of an Effective Assessment Tool

When evaluating an assessment tool, it is important to look at the characteristics, and qualities that make it an effective means of assessing competency. At a minimum, there must be evidence to demonstrate the instrument's validity and reliability (3). An ideal assessment tool would also possess the following qualities as outlined by Turnbull et al.: flexibility, comprehensiveness, feasibility, timeliness, accountability and relevance (53).

i. Validity

Validity refers to the concept that the test “measures what it purports to measure” (56). While this may sound like a simple concept, it is actually quite complex. There are multiple aspects of validity, with a number of subjective and objective approaches described to help determine the overall validity of an assessment tool. These include face, content, construct, criterion and discriminate validity.

Face validity is a highly subjective form of validation that is used during the initial development phases of a test (56). It is essentially the degree to which a test seems reasonable to experts in the area (3). A number of experts review the test, and decide whether or not it appears to be an appropriate measure of the qualities being assessed.

Content validity is also a largely subjective form of validation. It is an estimate of how well the assessment tool encompasses all the relevant domains, based on a detailed examination of the contents of the assessment tool (56).

Construct validity is established by demonstrating agreement between a theoretical concept and specific assessment tool or instrument (57). For example, to demonstrate that a new surgical simulator has construct validity, expert surgeons should obtain higher scores than novice trainees performing the same tasks.

Criterion validity examines the degree to which a test or assessment tool correlates to other measures of performance. There are two types of criterion validity:

concurrent and predictive. Concurrent validity refers to the correlation or agreement between the scores of the assessment tool and the established "Gold Standard". This would be used to evaluate a new assessment tool or test that was being established to replace a current standard test (15). Predictive validity is defined as "the extent to which the scores on a test are predictive of actual performance" (58). An example of a test with high predictive validity would be a surgical skills simulator whose scores correlate directly to performance in the operating room. While each form of validity is useful, predictive validity is the most likely to provide a clinically meaningful assessment (56).

Discriminate validity describes the translation of simulator skill to skill in the operating room (59). It is an evaluation of whether or not the assessment tool generates scores that actually correlate with the appropriate factors (56). This requires that the factors that should correlate highly actually do, and the factors that should correlate poorly do correlate poorly. An example of an assessment tool with strong discriminate validity would be an evaluation that could differentiate ability levels within a group with similar experience, such as discriminating abilities between all of the residents in a postgraduate year one surgery program (56).

ii. Reliability

Reliability refers to the precision, consistency and reproducibility of a test. It is the extent to which a test will produce the same results when used many times under the same conditions (15). Reliability is measured as a coefficient on a scale from 0 to 1. A test with a reliability coefficient of 0 is completely unreliable. That is, the variability in test results is independent of candidate ability and may be completely attributed to error. A test with a coefficient of 1 indicates complete reliability. This is an unrealistic result that is rarely achieved. There is general agreement that any high stakes examinations or tests on which important decisions are to be based should have a reliability ≥ 0.80 (52, 60).

As with validity, there are a number of methods available to evaluate the reliability of a test or assessment tool. These include test-retest, split-half, equivalent forms, and inter-rater reliability.

The test-retest method determines the reliability of the test by administering it on two or more occasions. The score on the initial test is compared to the score on the repeat test. This method is often not practical. It can be difficult to administer the same test to trainees on more than one occasion. If the retest is administered too soon after the initial test, the examinees may remember the answers from the initial test. As well, it is often not possible to control for the information learned in between the assessments, particularly if the interval is long (52).

For the split-half method, test items from a single assessment are divided in half and an individual's performance on each are correlated. A similar but more complex method determines the internal consistency of a test via Cronbach's alpha coefficient. This method correlates the performance of individuals by determining all possible split halves (52, 56).

Equivalent forms reliability is used to test the consistency of the results of two tests constructed in the same way from the same content domain. In essence, two separate tests that address the same objectives are created. The results of the two tests are then correlated to get the equivalent forms reliability (61). This method can be labour intensive, as essentially two equivalent assessments must be created.

Inter-rater reliability is determined by having two or more assessors evaluate the same trainee's performance and then correlating the results (3). A test that has high inter-rater reliability demonstrates that there is a high degree of agreement between the raters. The test scores are reflective of the trainee's performance, and not the particular examiner administering the test.

iii. Flexibility

The assessment tool must be able to evaluate clinical competency across a wide spectrum of practice. It should be applicable in multiple domains, and be able to be used many times (52, 53). For example, the assessment tool can be used in both major and minor surgery, the ambulatory clinic and the inpatient setting.

iv. Comprehensiveness

An effective assessment will cover all relevant objectives and assess all corresponding aspects of a trainee's performance.

v. Feasibility

To be successful, an assessment should be practical, portable, relatively inexpensive or cost-effective and acceptable to both the examiner and the trainee. There are however, certain situations where a more labour intensive and costly assessment may be acceptable. An example of this would be a licensure examination (52).

vi. Timeliness

In order to ensure that an evaluation is accurate, it must be completed as close in time to the target behavior as possible. The longer an evaluation is delayed the less accurate it becomes, as recall of events degrades and becomes more subjective. For these same reasons feedback to learners is also more effective the closer it occurs to the interaction (52, 53).

vii. Accountability

The assessment technique must be accountable to all parties involved. This includes the program and institution, students, educators, licensing authority and the community. Assessments must be defensible and capable of providing a justifiable analysis or explanation of results (62).

viii. Relevance

The assessment technique must be seen as important and relevant to both the examiners and the examinees. Both negative and positive evaluation results must be acted upon. Negative evaluations should result in remediation where necessary and positive evaluations should favorably affect advancement. If an evaluation is not viewed as important it will not be an effective assessment tool (63).

D. How will the assessment be categorized and judged?

In general there are two categories of both assessment and judging: formative and summative assessment, and norm or criterion-referenced judging.

In formative assessment, the evaluation is intended to provide constructive feedback to the learner during their training. Findings from a number of relevant assessments are combined and any strengths and weaknesses are addressed with the trainee. Any need for remediation may be highlighted, and objectives for subsequent assessments can be created. The focus of this type of assessment is on the individual, and not their ranking within their peer group (52). In program formative assessment, the evaluation is designed to improve the overall quality of the program. In both types of formative assessment the evaluation is not intended to make a pass or fail decision (52).

In contrast, the findings of a summative assessment are designed to make a pass/fail decision. In the case of trainee summative assessment, all relevant information is collected and evaluated. A decision of whether or not the trainee should continue to the next level is then made (52). In this type of assessment, an individual is often compared to their fellow peers. In program summative assessment, the evaluation is used to decide if the program meets the accepted standards for the purpose of continuing, restructuring or terminating the program (52).

Norm and criterion-referenced tests are two common methods of judging the results of an assessment. Norm-referenced testing is the more conventional method and is used to describe a candidate's performance in terms of their position in a group (52). The number of students that will pass or be given a particular grade is predetermined, and individual results are usually reported as a percentage of correct responses. In contrast, in criterion-referenced testing the success or deficiency of an individual student's achievement is based on the proportion of prescribed criteria or objectives that the student has met (52). Therefore criterion-referenced testing has particular importance in professional education where there is more of a concern that the student attains a minimal level of competence rather than focusing on their ranking within a peer group. Neither form of testing is superior to the other for all educational purposes, and both have merit in different circumstances (52).

E. Who Should Perform the Assessment?

Traditionally in medical education, assessments have primarily been performed in a physician-assessing physician manner. This encompasses several options including peer assessment, self-assessment, and faculty evaluation of trainees. Non-surgeon assessors such as nurses, standardized patients, and anesthetists are also an option. These have become a larger part of the evaluation process over the years (52).

The most commonly utilized method is physician-assessing physician. The support for this approach is that experts in a particular area are better able than non-experts to understand the underlying nuances of their field. Support for this is demonstrated with the use of global rating scales, which require the user to have an understanding of the content domain being examined. These have been shown to be a more reliable form of evaluation than checklists, which do not require the user to possess the same expert knowledge (64).

In medical and surgical practice, ongoing education is considered a professional requirement. This is driven by a physician's self-assessment of their individual learning needs. However, research into the accuracy of self-assessment in higher education has had mixed results. A meta-analysis of quantitative self-assessment studies in higher education showed that students are poor to moderate judges of their own performance (65). A literature review by Gordon (66) led to the same conclusion. A 2005 study by Mandel et al. however, demonstrated that surgical residents were actually fairly accurate at assessing their technical skills (67). In this study, 92 surgical residents performed self-assessments on a 6-station bench model including open and laparoscopic skills. The residents task specific, overall, and global assessments correlated significantly with faculty evaluations (67).

It is important to remember a number of things when reviewing the self-assessment literature. The ability of an individual to assess and self-assess is not an innate skill. It is something that needs to be taught and learned over time. As well, in order for learners to efficaciously use assessment tools, they must be properly instructed in their use. Additionally, these assessment tools must be well validated and rigorously tested.

Peer assessment refers to the assessment of an individual's performance by another individual in the same field, with the goal to maintain or enhance the quality of the work or performance in that field (68). Peer assessment can be a valuable tool in medical education (55, 69). It can be used to stimulate trainee participation in educational activities, clarify assessment criteria, and improve team performance and individual effort. A systemic review by Secomb (70) in 2008 showed evidence for peer teaching as an effective evaluation tool, as did a review by Speyer et al. (54) in 2011. Caution does need to be exercised with regards to the assessment instrument used; validity and reliability need to be demonstrated.

Non-surgeons have been increasingly utilized as assessors. Standardized patients have been shown to be as reliable as physicians in the assessment of objective structured clinical exams (OSCE) after proper instruction (71). Several studies have also demonstrated the high reliability of nurse evaluation of humanistic qualities (71-73). These assessments mainly rely on checklists, and whether non-surgeon evaluators would be able to assess technical skill has not been established.

F. When should assessment occur?

Assessment can occur either at different points throughout a course or training, or during different stages of motor skill acquisition.

In 1967, Fitts and Posner described a three-stage theory of motor skill acquisition that is widely accepted in both the motor skills and surgical literature (74, 75). This model separates the acquisition of motor skills into three separate stages – cognition, integration and automation. In the cognition stage, the goal is for the learner to understand the task. Performance is erratic at this point and the procedure is carried out in distinct steps. An example of this is learning to place a suture and tie a surgical knot. During the cognitive stage of learning this skill, the learner must understand the basic mechanics of the skill - how to hold the needle driver, pick up the tissue, hold the tie and place throws, and how to move their hands. With practice and feedback, the learner reaches the integrative stage. Here, the learner is able to translate the knowledge learned into more appropriate motor behavior. Performance is more fluid, and the procedure is carried out with fewer interruptions. The learner is still thinking about how to move their hands and how much tension is required to pick up the tissue and secure the knot, but they are able to execute the task more fluidly with fewer interruptions. Finally, with more practice, the learner reaches the autonomous stage of motor skill acquisition. The learner is able to perform the task with speed, efficiency and precision. They no longer need to think about how to execute the particular task, and can concentrate on refining performance and other aspects of the procedure (74, 75).

Another way of looking at the timing of assessment is at various points throughout a course or program. The assessment can occur before or at the beginning of the course as a pretest, during the course as continuous assessment, or at the end of the program as a final assessment (54).

The pretest assessment is useful as it demonstrates the students pre-existing knowledge. The focus of the course may be adjusted accordingly; less time may be spent on areas that the learners are already knowledgeable in, and more on areas in which they are weaker. Pre-reading material can also be recommended.

Continuous assessment is useful as it demonstrates the learner's progress. Areas requiring remediation may be identified and the necessary steps taken before the final assessment.

End of course or final evaluation affords measurement of the degree to which overall course or program objectives have been met. This provides a final assessment to the trainee.

G. Current Assessment of Technical Skills in Surgical Programs

A trainee's technical skills are constantly evaluated throughout residency training (76). This is most often a subjective assessment performed by practicing staff surgeons as a resident rotates through a particular surgical service (40). At periodic intervals throughout the rotation, the staff surgeon fills out an in-training evaluation report (ITER) that is used to assess all areas of the trainee's surgical competence. The ITER is composed of a list of rating scales with sections such as knowledge, professionalism, technical skill, teamwork, and communication skills that the evaluator completes to evaluate the candidate's clinical competency (3). The purpose of the ITER is to provide an accurate assessment of the trainee's performance abilities and provide useful feedback to both the trainee and the program (53).

The ITER has some advantages as an assessment tool. It can encompass aspects of competency such as professionalism, teamwork and communication skills that may not otherwise be evaluated. It is inexpensive and no special equipment is required. As well it allows for multiple observations over time (3).

However there are criticisms of the ITER as a way to evaluate technical skills. ITERs have been found to have poor validity and limited inter-rater and intra-rater reliability (3, 53, 76). According to Turnbull et al, this is primarily due to two major factors (53). First, faculty evaluators often do not receive training in the evaluation process and correct use of the ITER, which results in a number of rater errors. Errors of distribution include a range restriction or central tendency error and a leniency/severity or dove/hawk error. In central tendency errors evaluators use the mid-range average values and fail to use the whole scale, while in leniency/severity error there is a tendency to assign predominantly low or high scores (77). Additionally, ITERS may be subject to the halo effect, a correlation error that commonly occurs where the raters overall impression of the trainee influences the assessment of each individual component of the evaluation (77). Second, ITERs are not usually filled out on a daily basis, but rather in a retrospective fashion at the middle or end point of a rotation. Consequently, they are based on a recall of events and are generally not sufficiently detailed, or accurate.

Procedural logs record the number and type of procedure that a resident completes throughout their training and are often used as a marker of technical competence (76). These are a requirement for specialty certification by the American Board of Surgery, where an applicant must demonstrate their participation in a minimum number of cases in the various areas of general surgery (3). The problem with this is that the number of cases required to achieve competency can vary substantially between individuals, and participation in a procedure does not ensure competency (78). As well, the reliability of the data recorded can potentially be questionable. Trainees are often in a rush while logging procedures, or do so retrospectively at a time remote from the operation.

Currently, the majority of surgical programs across the country use a combination of ITERs and procedural logs to assess whether or not a trainee's technical skills are sufficient to progress to the next level in their training.

ITERs are summative as opposed to formative measures of assessment. They do not provide residents with a concrete understanding of the particular aspects of their technical skills that need improvement (79). On the other hand, global rating scales are reliable, have high inter-rater reliability and construct validity (80, 81). Residency programs across the country are trying to incorporate the use of these instruments into the evaluation of their trainee's technical skills. However this does not appear to be being done routinely in the majority of programs.

Chapter 3 Currently Validated Instruments Available for Evaluation of Surgical Technical Skills

A. Objective Structured Assessment of Technical Skills (OSATS)

In 1997, Martin et al., developed OSATS as a new way to more objectively evaluate the technical skills of surgical residents (82). It is a performance-based examination designed to assess the technical competency of surgical residents that consists of eight 15-minute bench model stations. Each model simulates portions of operative procedures relevant to general surgery. Examples of these stations include closure of a skin incision, insertion of a T-tube, hand-sewn bowel anastomosis, and control of an inferior vena caval hemorrhage (83). Two or more observers evaluate each task using a global rating scale with seven dimensions that are each related to some aspect of operative performance. The dimensions assessed include respect for tissue, time and motion, instrument handling, knowledge of instruments, use of assistants, flow of operation and forward planning and knowledge of specific procedure (82). Each dimension is graded on a five-point scale with the middle and extreme points anchored by specific descriptors to help clarify scoring (83). There is also an overall pass/fail judgment for the examiner to make based on their perception of overall performance (82). In this original study and a subsequent follow up of larger size, the global rating scale was compared to task specific checklists. Both studies demonstrated stronger construct validity and greater reliability for the global rating scale compared to the checklists, and the checklists are no longer used as part of the OSATS scoring assessment.

The OSATS (Appendix 1) has been extensively validated and shown to be an effective method for assessment of surgical skills (84). It does however require significant resources and time on the part of faculty evaluators (76). Over the past fifteen years, it has moved from use in the laboratory setting to use in the operating theatre (31, 50, 60, 80, 81). OSATS is validated and used for the evaluation of both open and laparoscopic surgical skills. It is viewed as the standard for skills assessment (50, 60). For these reasons OSATS was one of the evaluation tools used in this study.

B. Global Operative Assessment of Laparoscopic Skills (GOALS)

Originally developed by Vassiliou et al., GOALS is a global assessment tool used to assess intraoperative skills in laparoscopy (79). It is composed of a five item global rating scale and two visual analog scales (VAS) for overall competence and case difficulty. The five domains examined are depth perception, bimanual dexterity, efficiency, tissue handling, and autonomy (79). Each domain is scored on a five-point scale for dissection of the vesicular bed during laparoscopic cholecystectomy, with the middle and extreme points anchored by specific descriptors to help clarify scoring (85). The two VAS are 10 cm lines with descriptive anchors at each end. The first VAS measures how technically difficult the operation is, and the second measures the operator's overall competence. The evaluator marks the appropriate spot on the VAS with an X (79).

Vassiliou and colleagues demonstrated construct validity not only on the overall score, but also for each of the five GOALS items individually (79). Excellent internal consistency and interrater reliability was also demonstrated for the five item global rating scale. Although the VAS did demonstrate construct validity, interrater reliability was weak and we feel that it does not add benefit to the five item global rating scale. It is still included at the bottom of the rating scale, however most studies published since the original in 2004 do not utilize the VAS.

Fried et al. further validated GOALS as an assessment tool by demonstrating that it has concurrent validity by comparing GOALS scores with assessment of skills using a simulator (15). In 2007, Gumbs and colleagues demonstrated construct validity for GOALS for assessment of a complete laparoscopic cholecystectomy and laparoscopic appendectomy (85). This has added further evidence to the value and effectiveness of GOALS as a global assessment tool for laparoscopic skills.

To date GOALS is the only global rating scale that has been developed and validated specifically for the assessment of laparoscopic skills. For this reason, GOALS was included as one of the assessment tools in this study. (Appendix 2).

C. McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS)

In 1998, Derossis et al., set about developing a method to quantitatively assess technical skills in laparoscopic surgery by measuring performance through a series of exercises performed on a surgical simulator (8). The result was the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) simulation system. The original program consisted of seven standardized tasks performed in a trainer box under videoscopic guidance. The tasks were derived by a review of videotapes of basic and advanced laparoscopic procedures by expert laparoscopic surgeons. The face and content validity of the tasks was confirmed through questioning a large group of 44 expert laparoscopic surgeons (15). The original seven tasks ranged in difficulty and the skills they aimed at developing. Some of the exercises were chosen to develop basic dexterity skills, whereas others emphasized the use of specialized instruments and particular laparoscopic techniques. The tasks consisted of peg transfer, pattern cutting, clip application, placement of a ligating loop, mesh placement over a defect and intracorporeal and extracorporeal suturing (8). Since the original publication by Derossis et al., the clipping and mesh placement tasks have been eliminated secondary to a lack of validity and limitations in feasibility and cost (3).

Performance of each task is graded objectively, taking into account both precision of performance and speed. The score for an individual task is calculated by subtracting the time required to complete the task, in seconds, from a preset cutoff time (40). If the time to complete the task surpassed the preset cutoff time, a score of 0 was assigned. A penalty score for inaccuracy is also deducted according to a predetermined system (8).

The MISTELS system has been further validated since the original study and has been shown to be a highly reliable and valid system (3, 9, 15, 23, 40). In 2004, Fried et al., demonstrated predictive validity of the MISTELS model by showing that its scores correlate highly with intraoperative measurements of technical skill in laparoscopy ($r=0.81$; $P <0.001$) (15). For this, they evaluated a laparoscopic cholecystectomy using a modified version of Vassiliou et al.'s GOALS rating scale (15). The Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) have made MISTELS the manual skills component of its Fundamentals of Laparoscopic Surgery (FLS) program. These skills are the current gold standard for the assessment of laparoscopic skills outside of the operating room. As such, they were used as the video trainer manual skills component for this study.

D. Fundamentals of Laparoscopic Surgery (FLS)

Fundamentals of Laparoscopic Surgery (FLS), is an educational program developed by SAGES. This program was launched in 2004 (35). The overall goal of the FLS program is to teach and assess the basic cognitive and psychomotor skills required to competently perform laparoscopic surgery (35).

The FLS program consists of a didactic module and manual skills training practicum for education, and an examination component to assess competency. The cognitive portion of the educational program addresses four broad content areas: preoperative, intraoperative, and postoperative considerations and basic laparoscopic procedures (36). The manual skills training portion of the program is based on the MISTELS program originally developed by Derossis and colleagues at McGill University. Competence is assessed through a two-part examination, consisting of a cognitive computer based test and a proctored technical skills component.

The FLS program has been thoroughly tested and validated as a teaching and assessment tool for laparoscopic skills (8, 15, 16, 34-43). It meets the standards for high stakes examination. As of 2010, the American Board of Surgery requires its candidates to be FLS certified (34, 44).

E. Imperial College Surgical Assessment Device (ICSAD)

The Imperial College Surgical Assessment Device (ICSAD) is another method of assessing surgical skills that utilizes hand motion analysis. Sensors are placed on the dorsal surface of the trainee's hands while they perform a task and the hand motion is tracked. Their movement is translated into a computerized tracing of hand motion by the sensors. This provides an index of technical skill in both laparoscopic and open procedures that has been shown to have good concordance with OSATS scores (75, 86). Although the ICSAD has shown to be a valid indicator of operative skill, the evaluation metrics that are produced do not provide the trainee with feedback that can be used effectively for educational purposes (15, 76).

Chapter 4 Potential Models Available for Surgical Skills Assessment

A. Patients

While there has been a move away from teaching basic skills in the operating theatre, patients remain an integral part of surgical training. An operation can be reduced into its component parts that can be learned and mastered over a number of operations on separate patients. This is best done under the direct guidance and supervision of a staff surgeon (87). However, there are a multitude of ethical considerations when allowing trainees to operate on patients. This, combined with fiscal restraints, work hour reductions and restricted operating room time, make it important to evaluate other models for their role in assessment of surgical skills.

B. Cadavers

Cadavers have traditionally been used for medical and surgical anatomy sessions throughout training programs (84). However preserved cadaveric tissue does not resemble living tissue and it not a good model for laparoscopic surgery. The tissues are much stiffer, less malleable and difficult to work with. Depending on the preserving agent, it is extremely difficult or impossible to create a pneumoperitoneum. Additionally, cadavers are expensive and in limited supply (87, 88).

C. Animals

Animal models have been used extensively in medical training in the past. There are moral and ethical issues surrounding the use of animal models. Issues such as: is it absolutely necessary to use the animal model, or could these skills be learned in another way? Are the animals undergoing any harm or suffering as a result? Do the benefits to the learner justify the use of the animal? As well, the high cost and significant resources required to use these animal models, have led many universities and programs away from their use (84).

D. Video Trainers

These refer to the low fidelity physical box trainers or video trainers. These physical box trainers are composed of frames supporting traditional laparoscopic video monitors, light sources and camera systems (30). Trainees use actual laparoscopic instruments to perform tasks under videoscopic guidance (3).

Advantages of the physical box trainers are that they are portable, reusable, and they use the same instruments as used in the operating room (75). They are also relatively inexpensive with the average cost of a video trainer set up being approximately $2000.

Disadvantages include the fact that they are a low fidelity technology, the tasks being performed are basic and it is not possible to perform whole operations. Being able to walk through the steps of an operation helps the trainee to automate the steps and to troubleshoot any difficulties they may encounter in a safe environment (75). Additionally, an evaluator needs to be present to ensure the tasks are being performed properly and to evaluate the trainee. This limits timing and scheduling and potentially adds an increased cost if the evaluator requires compensation for their time.

E. Virtual Reality Simulators

Virtual reality trainers use computer-generated environments to perform computer-generated tasks. Computer-generated images are linked to a human-computer interface enabling the trainee to manipulate the images and receive objective feedback on performance from the computer (89).

There is a spectrum of machines available now, with some of the more recent ones incorporating haptic feedback. In surgery haptic or force feedback refers to the sense of touch that a surgeon experiences, both consciously and unconsciously, while operating (90). Adding haptic feedback to virtual reality simulators used for laparoscopic training and assessment is thought to be beneficial to trainees, especially during the early phase of psychomotor skill acquisition and for complex tasks (90-93).

Advantages of the virtual reality simulators are that they are reusable, there is no need for an evaluator as there are preset evaluation metrics on the computer, and data can be captured and saved for review at a later time. There is the ability to perform whole procedures, not just technical skills tasks. For example a trainee can perform a laparoscopic cholecystectomy from start to finish. As well, the level of difficulty can be adjusted to create easier or more challenging/stressful situations (22, 75).

Disadvantages of the virtual reality simulators include maintenance, three dimensions are not always well simulated, haptic feedback often lacks realism, and acceptance by trainees is often low (75). As a result trainees focus and effort on the tasks may be minimal. It then becomes less effective as a training tool and is not an accurate representation of actual laparoscopic skill. They are also extremely expensive compared to the video trainers, with the average cost of a virtual simulator ranging from approximately $80,000 - $120,000. That does not include the maintenance fees, technical support, or the purchase of any additional modules for the machine.

Chapter 5 Comparison of Video Trainers to Virtual Simulators

Training in lab-based settings with either virtual reality simulators or video trainers results in improvement of skills both in the simulation setting and the operating room. Simulation is now established as an effective method of skill acquisition (15-33).

In the first study to validate the transfer of training skills on a virtual reality simulator to the operating room, Seymour et al. demonstrated improved operative performance during laparoscopic cholecystectomy for residents that had trained on the MIST VR simulator (18). They showed that residents who had trained on the virtual reality simulator made fewer errors and were faster at completing the cholecystectomy than the control group (18). They did not include a video training group in their study.

In 2000, Scott and colleagues demonstrated that 30 minutes of daily video training for 10 days improved video-eye hand skills and translated into improved operative performance for junior surgical residents (31). Second and third year general surgery residents were prospectively randomized to either a video training group or a control group. Both groups were evaluated performing a pre and posttest laparoscopic cholecystectomy using a validated global operative assessment tool by three evaluators blinded to the resident's randomization status. All residents were also evaluated performing five standardized video trainer tasks at the beginning and the end of the study. The trained group achieved significantly greater improvement in video trainer scores on all five video trainer tasks, and significantly improved on four out of eight criteria on the global assessment scores for the laparoscopic cholecystectomy (31).

The benefit of one simulation method over the other for training purposes has not been clearly established. Hamilton et al. randomized junior general surgery residents to either a virtual reality or video trainer structured training program, and assessed their baseline skills compared to their post training skills (30). There was no statistically significant improvement difference between the two groups, with the virtual reality group improving their posttest score by 54%, and the video trainer improving theirs by 55%. They also looked at the effect of practice on operative performance by assessing all second year residents on their operative performance during a laparoscopic cholecystectomy before and after skills training. Operative performance improved only in the virtual reality group ($p<0.05$) (30).

In another study comparing virtual reality training to video training, Munz et al. randomized 24 novices to a either a control group, video training, or training on the LapSim virtual reality simulator (17). The 16 participants assigned to one of the training groups each performed 30 min training sessions once per week under direct supervision. Analysis of pre and post-test scores showed significant improvement in both trained groups, but no significant difference in the improvement between the two groups (17).

A recent Cochrane systematic review concluded that virtual reality training was at least as effective as video training in supplementing standard laparoscopic training, but no clear advantage was demonstrated (19).

Both video trainers and virtual reality simulators have demonstrated significant correlations between operative performance and psychomotor performance in lab-based settings and can be used to assess operative laparoscopic skills (8, 15, 25, 34, 37, 39, 41-43, 45-50). There is however a significantly larger body of evidence supporting this for the video trainer than the virtual reality simulator, and video trainers are currently used for the technical skills assessment component in the FLS program.

In 2003, Gallagher et al. demonstrated construct validity of the MIST VR virtual reality simulator, and showed that it was capable of evaluating the psychomotor skills required for the performance of laparoscopic surgery (48).

A study published by Kundhal and Grantcharov in 2008 showed validity of the LapSim virtual reality simulator for assessing laparoscopic skill (25). They recruited ten surgical residents with limited laparoscopic experience and had them perform three repetitions of seven basic skills tasks on the LapSim trainer, and then one laparoscopic cholecystectomy in the operating room. Time, error, and economy of motion parameters were measured on the LapSim. Operative performance was video recorded and blindly assessed by two independent observers using a modified OSATS rating scale. The correlation between time, economy of motion, and error parameters during the simulated tasks on the LapSim and the laparoscopic cholecystectomy were statistically assessed. Significant correlations between operative performance and simulator performance were demonstrated (25).

A larger prospective study by Langelotz and colleagues out of Germany established construct validity of the LapSim virtual reality simulator and showed that it could be used to assess operative laparoscopic skill (46).

To our knowledge, there have been no studies directly comparing video trainer and virtual reality simulation for assessment of operative laparoscopic skills. This has implications for surgical training programs, locally, nationally and internationally. Currently, most programs use a combination of simulation technology for training and assessing their residents. If either video trainers or virtual reality simulators are shown to be superior at assessing laparoscopic operative skill, resources could be directed towards utilizing that modality for evaluation purposes. This is particularly important as there are ever-increasing time and fiscal constraints in surgical training.

Chapter 6 Study Objectives

The objectives of this study are to

1. Establish construct validity for FLS, LapVR, GOALS, and OSATS in a mid sized academic center
2. Determine predictive validity of FLS and the LapVR virtual reality simulator using a human cholecystectomy model

Chapter 7 Methods

A. Participants

General surgery residents in postgraduate years 1-5 (PGY 1-5) and urology residents in PGY-1 at the University of Manitoba were invited to participate in this study. Informed consent was obtained from all participating subjects (Appendix 3). Ethics approval was obtained from the University of Manitoba Health Research Ethics Board (HREB).

B. Sample Size

Formal sample size and power calculations were not possible due to this study not having an intervention and therefore no effect size. Instead the educational literature and guidelines pertaining to these types of studies were reviewed in order to determine an appropriate sample size. Twenty-six general surgery and urology resident’s participated in the study. This is consistent with related educational literature (25, 94-96) and the American College of Surgeons Division of Education guidelines for studies of this nature (57).

C. Study Design

Data collection occurred between September 2011 and January 2013. All invited residents were given an explanation of the study and a consent form to review. Residents then performed the manual skills portion of the FLS tasks on the video trainer, the specified tasks on the LapVR virtual reality simulator, and a human laparoscopic cholecystectomy. Half of the residents performed the video trainer tasks first, while the other half performed the virtual reality simulator tasks first to control for any potential learning crossover between the two simulation modalities. The order of performance of the laparoscopic cholecystectomy and lab-based simulation activities was not always the same, but were done as close in time as possible.

Before each session, every resident completed a questionnaire outlining their prior laparoscopic experience and any practice on video trainers or virtual reality simulators, as well as any interim practice between the sessions. (See Appendix 4 for a copy of the questionnaire).

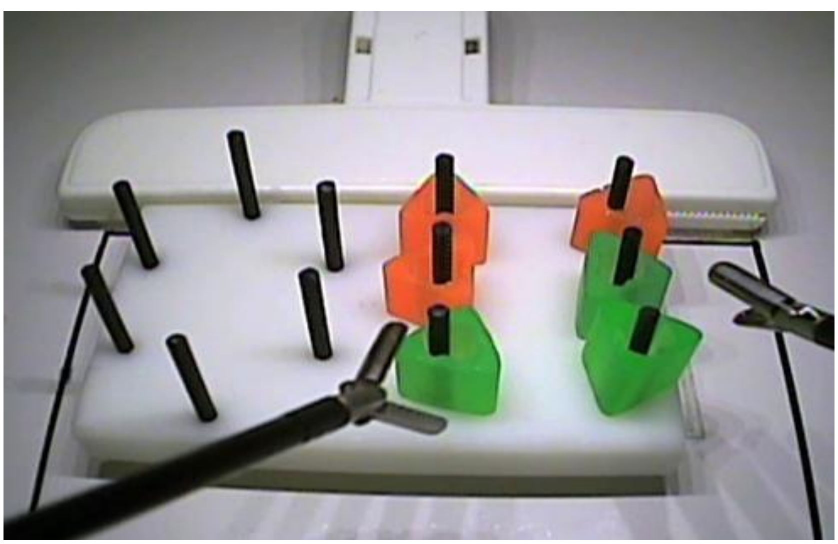
i. Box Trainer: Fundamentals of Laparoscopic Surgery

The manual skills portion of FLS consists of five tasks (8, 15). Participants performed four of these five tasks.

1. Peg Transfer

There are two pegboards and six pegs. The operator is required to use laparoscopic graspers to lift each peg with the left hand, transfer it to their right hand and place it on the second pegboard. Once this is completed with all six pegs, the sequence is reversed. The cutoff time is 300 seconds. A penalty score is given by calculating the percentage of pegs that could not be transferred as a result of being dropped outside the field of vision.

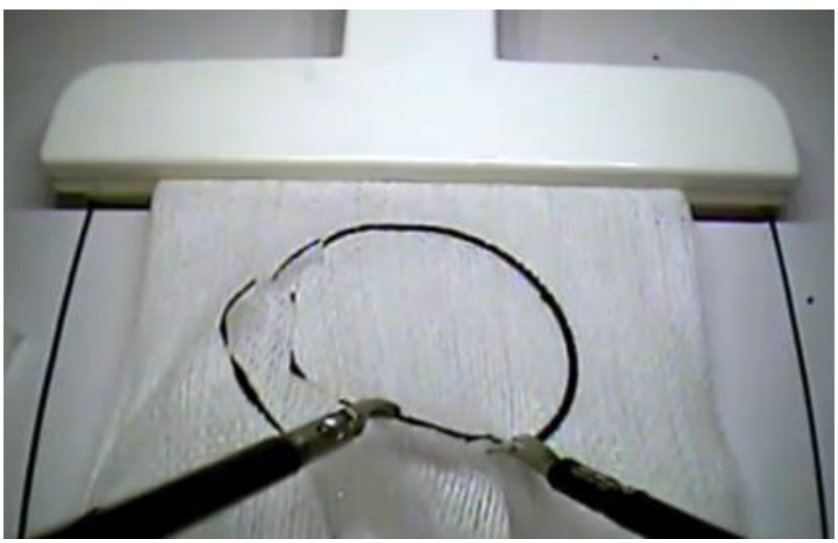
Figure 1. Picture of Box Trainer Peg Transfer Task



2. Pattern Cutting

For this task, the operator must cut out a pre-marked circular pattern from a suspended 10 x 10 cm gauze sheet. The grasper in the non-dominant hand places the gauze under tension while the operator uses their dominant hand to cut out the pattern with endoscopic scissors. The cutoff time is 300 seconds. A penalty score is determined by calculating the percentage of the area of deviation from the original pattern. For simplicity we deducted a standard penalty of 75 points for any deviation greater than 2 mm from the pre-marked pattern (97).

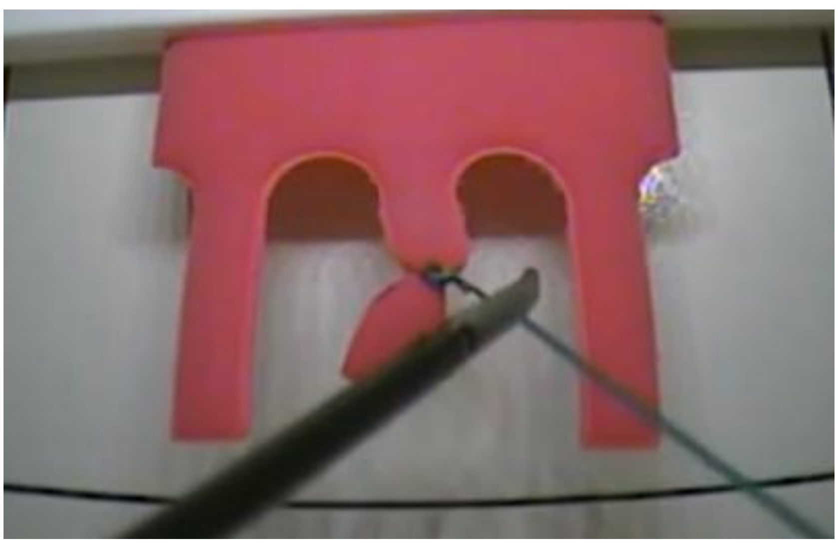
Figure 2. Picture of Box Trainer Pattern Cutting Task



3. Ligating Loop Application

This task requires placing a pre-tied slipknot at a specific pre-marked point on a foam tubular appendage. The operator uses a laparoscopic grasper to stabilize the foam appendage, while using their opposite hand to place the loop on the target, secure the knot, and then cut the excess suture. The cutoff time for this task is 180 seconds. The distance in millimeters that the knot deviates from the pre-marked line determines the penalty. An additional deduction of 50 points is added to the penalty score for insecure or failed knots.

Figure 3. Picture of Box Trainer Ligating Loop task



4. Intracorporeal (IC) Suturing

This task requires the operator to place a simple suture (2-0 silk suture 12 cm in length on a curved v-20 needle) through pre-marked points on a longitudinally slit Penrose (rubber) drain. The operator then creates three knots using an intracorporeal knot tying technique. The needle is passed to the opposite hand with each knot. The cut-off time is 600 seconds. The penalty score is the sum of the distance in millimeters that the suture placement deviates from the pre-marked points, plus the gap in millimeters between the edges of the approximated Penrose drain. An additional penalty is applied for knots that slip (10 points) or come apart (20 points).

Figure 4. Picture of Box Trainer Intracorporeal Suturing Task

Didactic instruction and demonstration were given to each participant on how to perform the tasks. All subjects performed one repetition of the four tasks consecutively as a familiarization trial. They then repeated the sequence a second time, with this data being collected for analysis purposes.

ii. Virtual Reality: LapVR

The LapVR platform was used for the virtual simulator component of the study. It is a personal computer based virtual reality system consisting of a 20-inch flat liquid-crystal digital display, two laparoscopic instrument handles with options for needle driver handles, a camera device and two foot pedals. The software consists of four different modules: essential skill, procedural skill, obstetrics/gynecology module, and general surgery procedures module. It also contains an original haptic feedback system (98).

The essential skill module contains six tasks: camera navigation, peg transfer, pattern cutting, clipping, needle driving, and knot tying at three different levels (1, 2 and 3 representing beginner, intermediate, and advanced levels respectively). The LapVR automatically evaluates each task by summing a total score of 100 parameters, which include factors such as time taken to complete the task, number of errors – e.g. extent of tissue damage, number of wrongly placed clips, estimated amount of blood loss, etc., and instrument path length.

Peg transfer, pattern cutting, clipping, needle driving, and knot tying at beginner level 1 were included for the virtual simulator portion of the study. These were felt to most closely approximate the FLS video trainer tasks outlined above.

Older virtual simulators such as the MIST-VR have tasks that are completely unrelated to the FLS video trainer tasks. For example, the MIST-VR has a number of tasks that involve touching, grasping and cauterizing virtual spheres and plates within a three-dimensional cube (49).

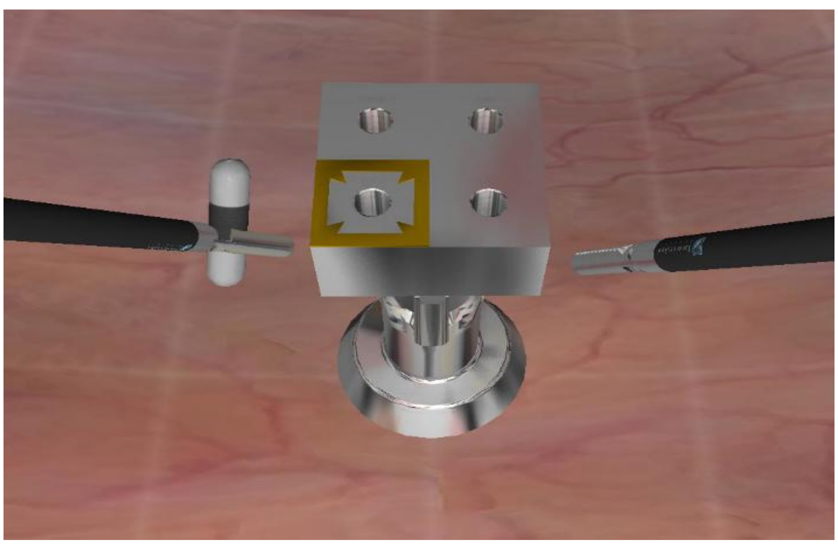
Newer devices are including tasks that are more similar to the MISTELS video trainer tasks as well as whole operations such as laparoscopic cholecystectomies.

LapVR Tasks:

1. Peg Transfer

There is one pegboard with four holes present in the center of the screen. The user must pick up pegs that appear on alternating sides of the screen and place them in the highlighted spot on the pegboard. By having pegs appear on alternate sides of the screen, the user must use both the dominant and non-dominant hands approximately equally to complete the task. The cut-off time is 360 seconds.

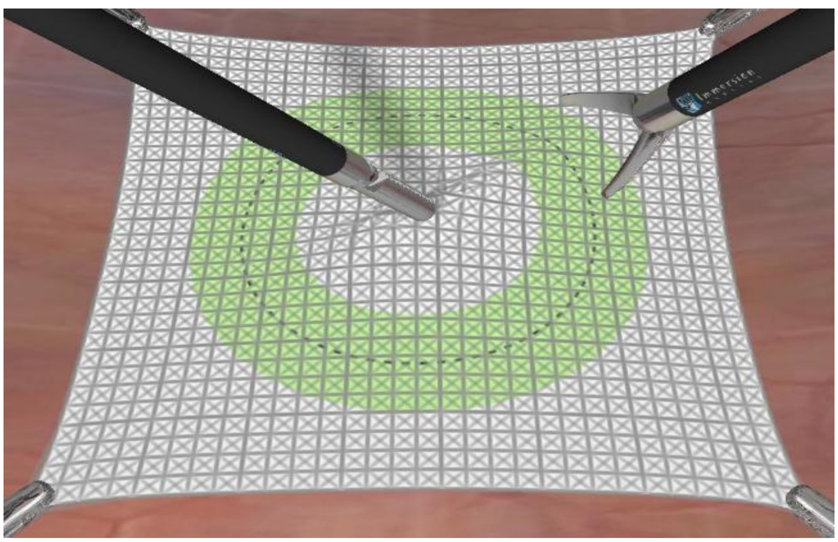
Figure 5. Picture of LapVR Peg Transfer Task



2. Pattern Cutting

This task is very similar to the pattern-cutting task in the manual skills portion of FLS. The operator must cut out a pre-marked circular pattern from a suspended gauze sheet. The cut-off time for this task is 300 seconds.

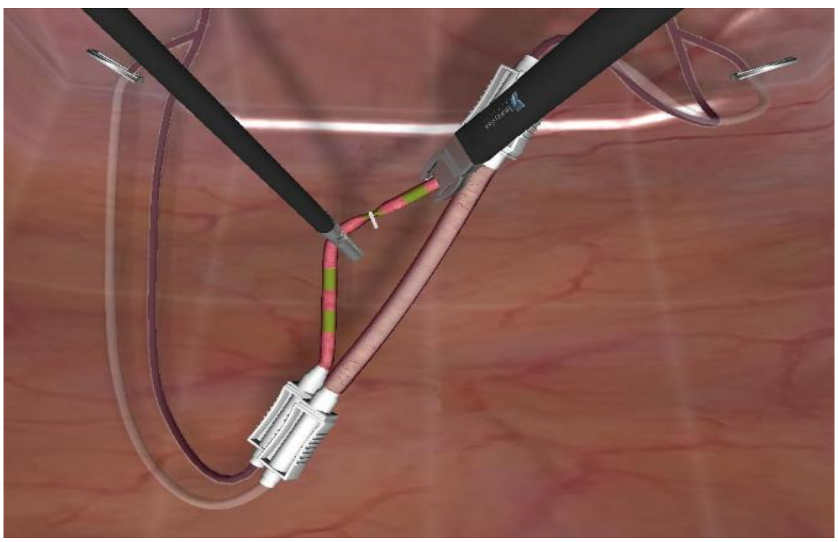
Figure 6. Picture of LapVR Pattern Cutting Task



3. Clip Application

There is a representation of an artery, and a ductal structure in the center of the screen. The ductal structure has four highlighted areas where the operator must apply clips. Once clips are applied to the correct areas, a highlighted mark appears in the center where the duct is to be divided. The cut-off time is 180 seconds.

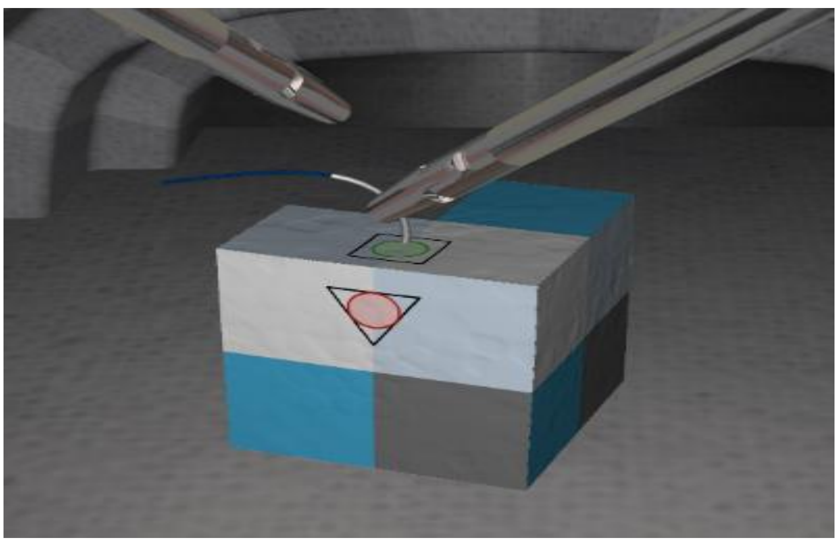
Figure 7. Picture of LapVR Clip Application Task



4. Needle Driving

For this task there is a series of four separate blocks that appear sequentially in the middle of the monitor. Each block has a pre-marked area where the operator must drive a needle through and remove it from the corresponding designated area of the block. The cut-off time is 300 seconds.

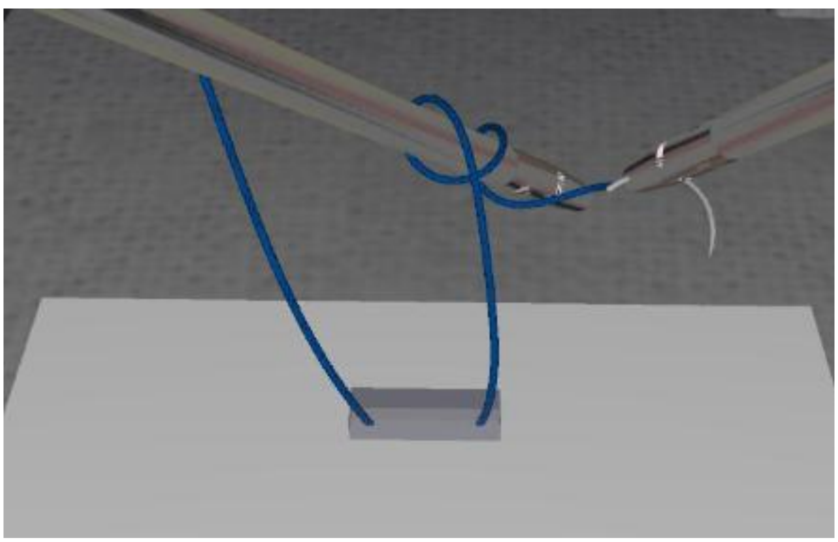
Figure 8. Picture of LapVR Needle Driving Task



5. Knot Tying

The operator is required to place three knots using an intracorporeal knot tying technique. The first knot is a double throw; the second and third knots are opposing single throws. The cut-off time for this task is 300 seconds.

Figure 9. Picture of LapVR Knot Tying Task



All subjects were given instruction and demonstration in how to perform the tasks using correct technique and instrument handling. They completed the five tasks consecutively as a familiarization trial, and then repeated them for assessment purposes and data collection.

iii. In Vivo: Human Cholecystectomy

Each resident performed an elective laparoscopic cholecystectomy on a human patient under the direct supervision of an attending surgeon. If the cholecystectomy was deemed to be too difficult for the level of the resident at the beginning of the operation, a new case was selected for evaluation. The staff surgeon used the previously validated GOALS assessment tool (15, 79, 85) and a modified version of Martin and Reznick's OSATS (82, 84) to evaluate the resident's operative performance. There were three staff surgeons involved with the study, all had been instructed and were very familiar with the use of the assessment tools. They were instructed not to help the resident or provide guidance unless the resident failed to progress appropriately, or if patient safety became a concern such as with significant bleeding or unusual anatomy.

D. Outcome Measures

Outcome measures for the separate modalities included: FLS scoring system metrics (8, 15, 37, 41), computer based evaluations (98), and operative evaluations based on global operative rating scales (15, 79, 82, 84, 85). All forms of these outcome measures have been previously validated.

For the FLS scoring system metrics, a cut-off time is assigned for each task (8, 15). A raw score is calculated by subtracting the subject's time taken to complete the task in seconds plus any additional penalties from the cut-off time, as follows:

Raw score = Cut-off time – Completion time – Penalty score

As a result, higher scores indicate superior performance. Each subject's raw score was then normalized to provide equal weighting of the tasks according to formula's available through the FLS program (41). A normalized score of 100 on any individual task indicates an excellent performance. For each subject, the normalized score on each of the four tasks was summed to give an overall test score.

The computer based performance measures are automatic evaluations performed by the LapVR. The computer records a number of parameters related to time, error, economy of motion, and overall score. Iwata et al. validated a number of these parameters in 2010 (98). Forty-four subjects were divided into three groups consisting of no previous laparoscopic experience, minimal laparoscopic experience and experienced laparoscopic surgeons. All subjects were compared performing the six essential tasks on the LapVR. Differences between the scores for each group of participants were statistically evaluated. Significant differences were detected between the groups for 16 of the 100 scoring parameters (98). Using this study as a guide, the data for these validated parameters was extracted and recorded for each subject. This ranged from one validated parameter for the clipping task to four for the peg transfer and pattern cutting tasks.

Operative evaluations of the laparoscopic cholecystectomies were performed using the previously validated global operative rating scales OSATS and GOALS. A higher overall score indicates superior performance on both of these scales.

E. Data Analysis

The analyses were carried out using Statistical Analysis System SAS 9.2 (SAS Institute, Cary, NC). All p-values were two tailed and a p-value <0.05 was considered statistically significant. Continuous variables (e.g., GOALS score) were summarized in the descriptive analyses using the mean and standard deviations. Categorical variables (e.g., comprehensive score) were summarized using the number (percent) in various categories. Comparisons between junior residents (PGY 1 and 2) and senior residents (PGY 3-5) were conducted using t tests or non-parametric Wilcoxon-Mann-Whitney Test for continuous variables, and chi-square tests or Fisher's exact test for categorical variables. Multivariable regression analysis was carried out to examine the association between outcome variables (GOALS score or OSATS score) and predictor variables (FLS overall score or LapVR data from each domain: Peg Transfer, Cutting, Clipping, Needle Driving, and Knot Tying). The R-squares from these multivariable regression analyses were used as a measure of how strongly FLS and LapVR evaluation metrics predict intra-operative performance. R-square values greater than 0.25 are considered a large effect, 0.09 – 0.25 a moderate effect, and less than 0.09 is considered to be a small effect.

Chapter 8 Results

A. Participant Characteristics

A total of 26 subjects (14 female, 12 male) participated in the study (Table 1). Participants included postgraduate training year (PGY) 1 (n=12), PGY 2 (n=4), PGY 3 (n=3), PGY 4 (n=6), and PGY 5 (n=1). General surgery (n=23) residents of all PGY levels were included and urology (n= 3) PGY 1 residents were included. 24 participants were right-handed and 2 were left-handed. Data was analyzed for the separate junior (PGY1 and 2) and senior (PGY 3, 4, and 5) groups. Half of the juniors and half of the seniors performed the FLS video trainer tasks before the LapVR virtual simulator. The order of the laparoscopic cholecystectomy and simulation activities varied, as they were limited by opportunity and scheduling. 17 residents performed the simulation tasks prior to the laparoscopic cholecystectomy and 9 performed the laparoscopic cholecystectomy first. This was evenly divided between junior and senior residents. All previous and interim experience was documented through a questionnaire. Time between activities was also recorded. Interim experience was found to be more useful than time between activities, as many of the PGY 1 and 2 residents were rotating on non-surgical rotations and therefore not participating in any operative or simulation activities between sessions. All 26 participants completed the study.

Table 1. Characteristics of Study Participants

	Overall (n=26)	Juniors (n=16)	Seniors (n=10)	Significance level
Age	29.7 (3.4)	29.3 (3.9)	30.5 (2.5)	0.19
Gender				0.42
F, n (%)	14 (53.9)	10 (62.5)	4 (40)	
M, n (%)	12 (46.1)	6 (37.5)	6 (60)	
Handedness				0.99
L, n (%)	2 (7.7)	1 (6.3)	1 (10)	
R, n (%)	24 (92.3)	15 (93.7)	9 (90)	
Home program				0.26
General Surgery, n (%)	23 (88.5)	13 (81.2)	10 (100)	
Urology, n (%)	3 (11.5)	3 (18.8)	0	
Laparoscopic experience (procedures as primary operator)	60.83 (83.51)	4.84 (9.07)	150.4 (68.90)	**<0.001**
Video/box trainer experience (hours)	10.05 (29.73)	10.51 (37.23)	9.3 (12.0)	0.09
Virtual simulator experience (hours)	0.34 (0.54)	0.39 (0.61)	0.25 (0.42)	0.69
Time between activities (days)	93.12 (113.7)	149.40 (112.9)	3.10 (4.15)	**<0.001**
Interim laparoscopic experience (procedures as primary operator)	1.69 (3.03)	2.13 (3.16)	1.0 (2.83)	**0.03**
Interim video/box trainer experience (hours)	0.65 (1.15)	1.06 (1.31)	0 (0)	-
Interim virtual simulator experience (hours)	0.08 (0.27)	0.13 (0.34)	0 (0)	-

Note: Values reported are mean (standard deviation) unless specified.

Participant characteristics were generally similar between the junior and senior groups (Table 1). There was no significant difference in age (p=0.19), gender (p=0.42), handedness (p=0.99), or home program (p=0.26) between the two groups. As expected, there was a significant difference between the number of previous laparoscopic procedures performed between the juniors and the seniors ($p<0.001$). This was expected, as senior residents should have performed a significantly larger number of procedures than the junior residents. There was no difference between the number of hours spent practicing on video trainers ($p=0.09$) and virtual simulators ($p=0.69$) for both the juniors and seniors. There was a significant difference in the number of days between activities ($p<0.001$), with the junior residents having a much longer time period between activities than the seniors (149.4 days vs. 3.1 days). There was also a significant difference in the number of interim laparoscopic procedures performed between the two groups ($p=0.03$), with the juniors performing an average of 2.13 procedures and the seniors an average of 1.0 procedures.

B. Descriptive Statistics of Outcome and Predictor Variables

Table 2. Descriptive Statistics of Outcome and Predictor Variables

	Overall (n=26)	Juniors (n=16)	Seniors (n=10)	Significance level
Outcome Variables				
GOALS	15.04 (5.20)	11.50 (2.71)	20.70 (2.21)	**<0.001**
OSATS	28.69 (9.07)	22.38 (4.16)	38.80 (3.88)	**<0.001**
Predictor Variables				
FLS				
Peg Transfer	85.67 (12.4)	81.46 (11.4)	92.4 (11.4)	**0.03**
Pattern Cutting	16.79 (19.2)	10.3 (11.7)	27.1 (24.4)	0.07
Ligating Loop	68.47 (24.9)	58.6 (25.8)	84.3 (12.9)	**0.003**
IC Suturing	55.97 (33.0)	39.4 (28.2)	82.4 (20.7)	**<0.001**
Overall Score	226.9 (73.9)	189.8 (56.7)	286.3 (58.6)	**<0.001**

LapVR				
Peg Transfer				
Time (seconds)	130.4 (48.6)	139.6 (50.2)	115.6 (44.2)	0.23
Non-dominant Time (seconds)	35.92 (16.03)	39.23 (18.27)	30.0 (9.69)	0.14
Non-dominant Path Length	2.90 (0.95)	2.85 (1.01)	2.98 (0.87)	0.72
Comprehensive Score, Yes, n (%)	13 (50)	8 (50)	5 (50)	0.99
Cutting				
Time (seconds)	207.5 (146)	264.8 (160.0)	115.8 (36.4)	**0.008**
Unsuccessful Cuts	7.38 (7.68)	9.75 (8.76)	3.60 (3.17)	**0.02**
Non-dominant Path Length	2.01 (1.85)	2.42 (2.01)	1.38 (1.42)	0.17
Comprehensive Score, Yes, n (%)	16 (61.5)	9 (56.3)	7 (70)	0.48
Clipping				
Time (seconds)	74.85 (28.38)	89.13 (24.15)	52.0 (16.03)	**<0.001**
Needle Driving				
Time (seconds)	364.6 (173.8)	421.1 (176.1)	274.3 (132.3)	**0.03**
Dominant Path Length	7.13 (3.31)	8.38 (3.41)	5.14 (1.98)	**0.01**
Non-dominant Path Length	4.96 (2.97)	5.75 (3.01)	3.70 (2.56)	0.09
Knot Tying				
Time (seconds)	267.8 (170.5)	332.3 (163.1)	171.1 (137.4)	**0.02**
Dominant Path Length	6.39 (4.48)	8.04 (4.84)	3.91 (2.40)	**0.01**
Non-dominant Path Length	6.34 (4.65)	3.70 (2.56)	4.05 (2.70)	**0.04**

Note: Values reported are mean (standard deviation) unless specified.

Senior residents had significantly higher GOALS and OSATS scores than junior residents (20.7 vs. 11.5, p <0.001, 38.8 vs. 22.38, p <0.001) as seen in table 2. They also had significantly higher scores on the overall FLS and three out of the four FLS tasks (peg transfer: 92.4 vs. 81.46, p=0.03; ligating loop: 84.3 vs. 58.6, p=0.003; IC suturing: 82.4 vs. 39.4, p <0.001; overall score; 286.3 vs. 189.8, p <0.001), with the fifth task (pattern cutting) approaching significance (27.1 vs. 10.3, p=0.07). For the LapVR tasks there were no significant differences between the senior and junior residents for any of the peg transfer parameters. For the cutting task, the senior residents scored significantly better on time to completion (115.8 vs. 264.8, p=0.008), and unsuccessful cuts (3.6 vs. 9.75, p=0.02). The seniors also scored significantly better on the time to completion for clipping (52.0 vs. 89.13, p <0.001). For needle driving, time and dominant path length were statistically significant, with senior residents outperforming the junior residents (264.3 vs. 421.1, p=0.03; 5.14 vs. 8.38, p=0.01). The seniors had significantly better scores on the time to completion and dominant path length knot tying parameters (time: 171.1 vs. 332.3, p=0.02; dominant path length: 3.9 vs. 8.04, p=0.01), but the juniors scored significantly better on non-dominant path length (3.7 vs. 4.05, p=0.04).

C. Multivariable Regression Analysis for GOALS

Table 3. Multiple Correlations between GOALS scores and Predictors from Multiple Regression Analysis

	Overall (N=26)		Juniors (n=16)		Seniors (n=10)	
	R-square	p-value	R-square	p-value	R-square	p-value
FLS						
Overall Score	0.388	**<.001**	0.004	0.83	0.307	0.10
Peg Transfer	0.241	**0.01**	0.015	0.66	0.256	0.13
Pattern Cutting	0.243	**0.01**	0.015	0.65	0.195	0.20
Ligating Loop	0.210	**0.02**	0.001	0.89	0.064	0.48
IC Suturing	0.334	**0.002**	0.034	0.49	0.376	0.06
LapVR						
Peg Transfer	0.095	0.70	0.160	0.72	0.631	0.21
Cutting	0.315	0.08	0.334	0.30	0.111	0.95
Clipping	0.397	**<.001**	0.006	0.77	0.195	0.20
Needle Driving	0.241	0.10	0.116	0.67	0.211	0.67
Knot Tying	0.342	**0.03**	0.390	0.13	0.377	0.38

Overall FLS scores are positively and significantly associated with GOALS scores (R-square = 0.388, p <0.001, table 3). Additionally all four of the individual FLS tasks demonstrated significance. For the LapVR, clipping (R-square = 0.397, p<0.001), and knot tying (R-square = 0.342, p=0.03) are significantly associated with operative GOALS scores.

Figure 10. Relationship between overall FLS score and overall GOALS score

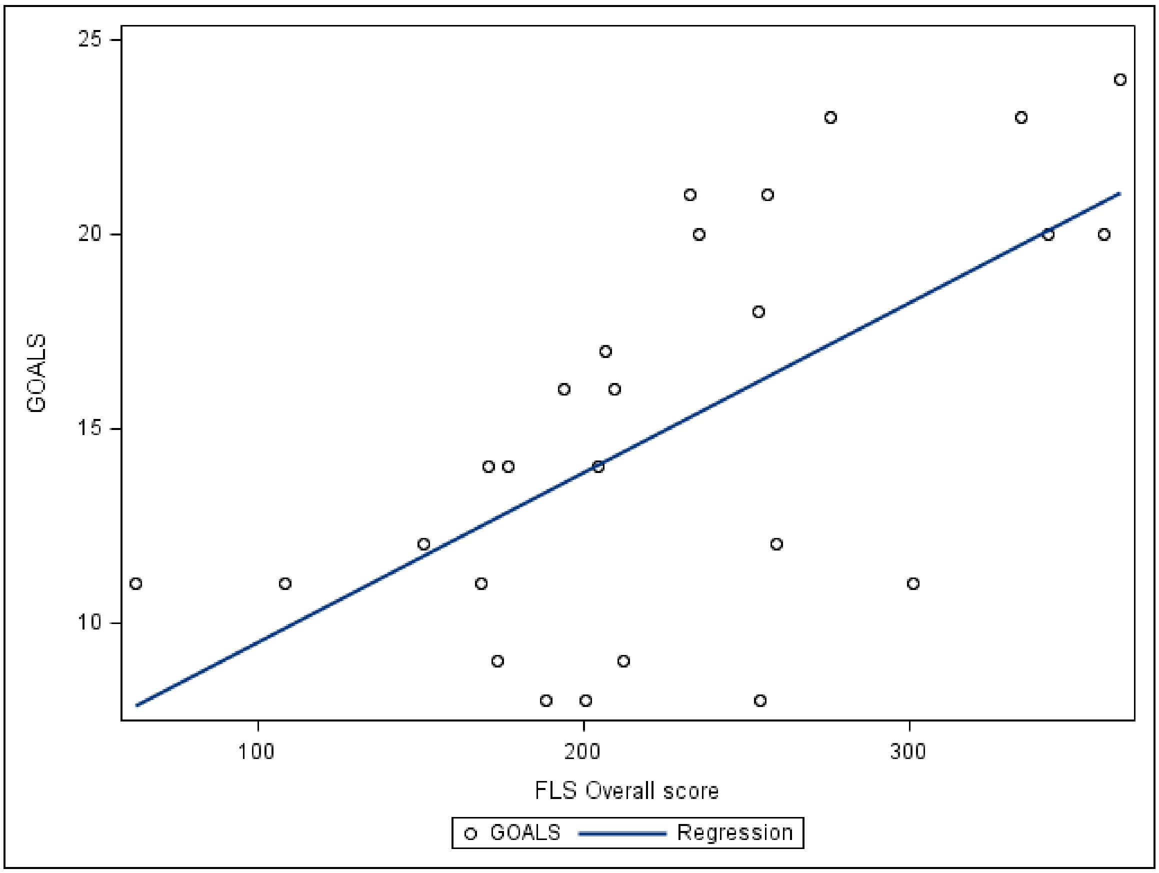


Figure 10 shows that FLS overall scores are positively associated with intra-operative laparoscopic cholecystectomy performance evaluated by GOALS assessment.

D. Multivariable Regression Analysis for OSATS

Table 4. Multiple Correlations between OSATS scores and Predictors from Multiple Regression Analysis

	Overall (N=26)		Juniors (n=16)		Seniors (n=10)	
	R-square	p-value	R-square	p-value	R-square	p-value
FLS						
Overall Score	0.415	**0.001**	0.002	0.88	0.334	0.08
Peg Transfer	0.254	**0.009**	0.019	0.61	0.283	0.11
Pattern Cutting	0.262	**0.008**	0.027	0.54	0.221	0.17
Ligating Loop	0.206	**0.02**	0.006	0.78	0.033	0.61
IC Suturing	0.377	<**0.001**	0.020	0.60	0.461	0.03
LapVR						
Peg Transfer	0.172	0.39	0.122	0.82	0.713	0.12
Cutting	0.399	**0.02**	0.339	0.29	0.247	0.80
Clipping	0.451	**<.001**	0.057	0.37	0.119	0.33
Needle Driving	0.257	0.08	0.052	0.88	0.213	0.67
Knot Tying	0.359	**0.02**	0.530	0.03	0.322	0.47

Table 4 shows that overall FLS scores are positively and significantly associated with OSATS scores (R-square = 0.415, p=0.001). All four of the individual FLS tasks also demonstrated significance. For the LapVR, cutting (R-square = 0.399, p=0.02), clipping (R-square = 0.451, $p<0.001$), and knot tying (R-square = 0.359, $p=0.02$) are significantly associated with operative OSATS scores.

Figure 11. Relationship between overall FLS score and overall OSATS score

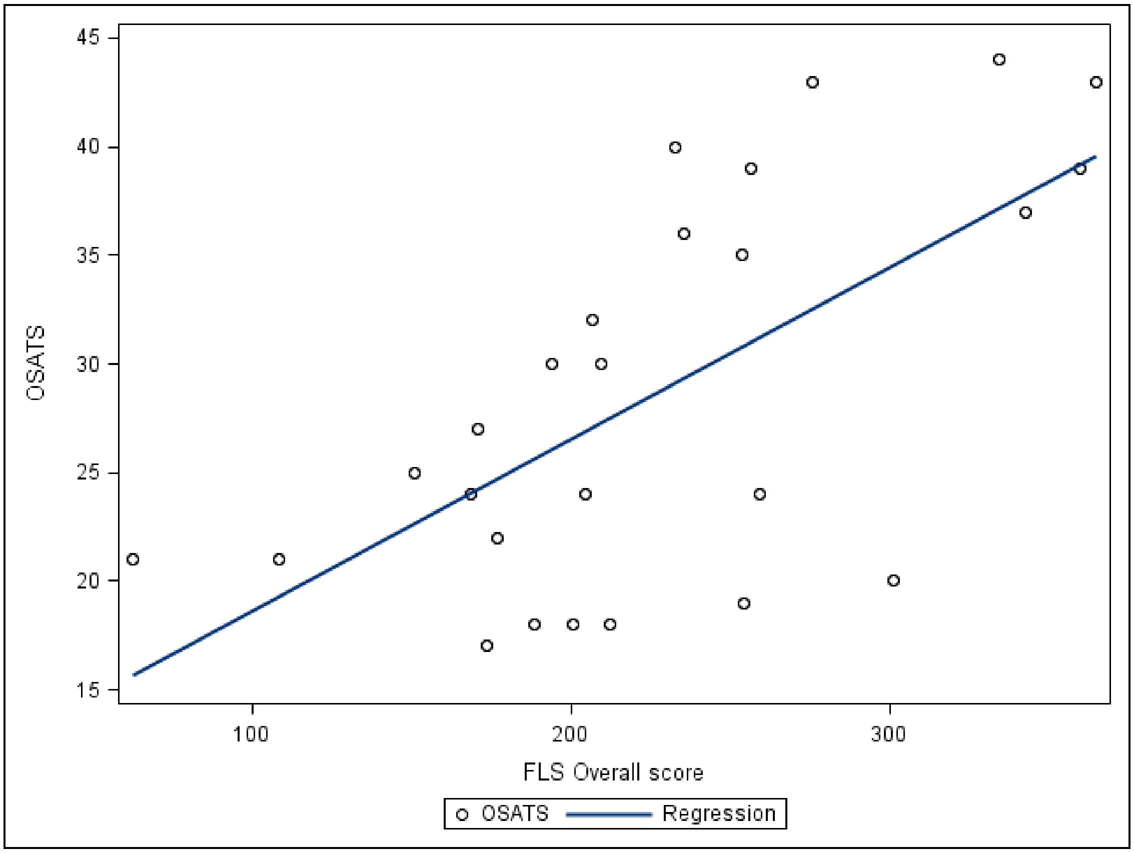


Figure 11 shows that FLS overall scores are positively associated with intra-operative laparoscopic cholecystectomy performance evaluated by OSATS assessment.

Chapter 9 Discussion

A useful framework to revisit when discussing the significance of video trainers versus virtual reality simulators for assessment of laparoscopic surgical skills, are the considerations that medical educators make when developing a method for assessing a learned skill (52-54). Namely: What are the attributes of an effective assessment? Why is the assessment being performed? What should be assessed? How will the assessment be categorized and judged? Who should perform the assessment? When should the assessment occur?

A. Requirements of an Effective Assessment Tool

An effective assessment tool must be both valid and reliable. Additionally, an ideal assessment tool would possess a number of other desirable attributes such as flexibility, comprehensiveness, feasibility, timeliness, accountability, and relevance.

i. Validity

a. Construct Validity

Both GOALS and OSATS score's were significantly different between the junior and senior groups for the laparoscopic cholecystectomy, with senior residents scoring significantly higher than the junior residents ($p<0.001$ for both, Table 2). This demonstrates construct validity and is consistent with the literature that evaluates these global rating scales (79, 84, 85).

The senior resident group also had significantly higher scores on the overall FLS score, peg transfer, ligating loop, and intracorporeal suturing. The only task that did not demonstrate a statistically significant difference was the pattern cutting. A possible reason for this could be the penalty scoring for this task. The penalty for cutting more than 2 mm outside of the target area was large at 75 points. Even if subjects completed the task within the designated time frame, the penalty for inaccuracy brought many of the subjects score down significantly. As overall FLS score was the measure we were comparing, construct validity was demonstrated for FLS.

For the LapVR tasks there were significant differences demonstrated between the junior and senior residents for all the clipping and knot tying parameters. However for one of the knot tying parameters, non-dominant path length, the junior residents had significantly better scores than the senior residents. This is not expected, as senior residents should have a more efficient path length than the juniors. A possible explanation is that junior residents may be ignoring their non-dominant hand, thus giving a falsely improved score. Differences were found for two of the three needle driving parameters and two of the four cutting parameters. None of the parameters evaluated for the peg transfer were significant. This is in contrast to the study validating the LapVR by Iwata et al. in 2010 (98). Possible reasons for this could be that the peg transfer task was the first task that the residents performed. As majority of the residents had spent minimal, if any time on a virtual simulator before, a lack of familiarity with the technology could be partially responsible. Secondly, at this institution, the peg transfer task seemed to be the task the junior resident's practiced most often on the video trainer. A crossover training effect between the video trainer and virtual simulator has been demonstrated (30), and this may have been partially responsible. Two of the parameters measured for the cutting task, non-dominant hand path length and comprehensive score, showed no significant difference between the junior and senior residents. This is again in contrast to previous findings by Iwata et al. (98). Junior residents would be expected to be less adept at skillfully manipulating their non-dominant hand than the seniors, and should have an overall lower success rate. This could be a result of poor performance by the senior residents and the sample size was not large enough to show a true effect. One of the validated needle driving parameters, non-dominant hand path length, also did not show a significant difference. This was the most difficult task according to participants, and many senior residents struggled to orient their instruments correctly. The lack of significance could also be related to sample size. Construct validity for the LapVR was completely demonstrated for one of the five tasks, partially demonstrated for three of the five, and not demonstrated for one of the five. Compared to FLS, construct validity was much weaker and less thoroughly demonstrated for the LapVR.

b. Predictive Validity

Overall FLS scores are significantly associated with both overall GOALS and OSATS scores (Table 3). Individuals with higher FLS scores are more likely to have higher GOALS and OSATS scores. This is consistent with previous literature (34, 36, 37, 39), and demonstrates both predictive and discriminate validity. The relationship is slightly stronger for FLS and OSATS, with overall FLS score explaining 41.5% of the variation in OSATS operative score compared to 38.8% of the variation in GOALS operative score. This however is not significant with a p value of 0.901. For Both GOALS and OSATS all of the individual FLS tasks also demonstrated a statistically significant relationship. The adds further strength to the predictive validity of overall FLS score.

The LapVR tasks had to be compared to overall operative score individually as currently there is no validated overall summary score for the tasks as in FLS. Additionally the individual LapVR tasks could not be compared to the individual FLS tasks as each LapVR task had anywhere from one to four scoring parameters contributing to the task evaluation. Looking at the relationship between the individual LapVR tasks and GOALS scores (Table 3), it can be seen that clipping, and knot tying are significantly associated with GOALS scores, with clipping score explaining 39.7% of the variation in GOALS operative score and knot tying explaining 34.2%. Therefore, predictive validity was only demonstrated for two of the five LapVR tasks and GOALS scores. Table 4 shows the relationship between individual LapVR tasks and OSATS scores. Clipping, knot tying, and cutting were statistically significant explaining 45.1%, 35.9%, and 39.9% of the variation in OSATS operative score respectively. Therefore, predictive validity was demonstrated for three of the five LapVR tasks and OSATS scores. Both predictive and discriminate validity have been demonstrated for virtual reality simulators, however the literature primarily looks at the MIST VR virtual simulator and the LapSim virtual reality simulator (25, 45-49). Prior to this, there have not been any studies looking at the

LapVR specifically. For this study, predictive validity was partially demonstrated for the LapVR, this is in contrast to FLS, where predictive validity was strongly demonstrated.

ii. Reliability

To use FLS and LapVR for assessment of surgical trainees prior to performing laparoscopic surgery on human patients there should be no doubt about the reliability of the assessment tool being used. Although we did not specifically look at reliability with this current study, there is information available on it in the literature.

The reliability of the FLS tasks has been well established (8, 15, 16, 34-43), and SAGES currently uses FLS for assessment of the laparoscopic skills component of their examinations. Fried et al demonstrated excellent interrater and test-retest reliabilities for overall MISTELS scores with intraclass correlation coefficients of 0.998 and 0.892 respectively (15). Cronbach's alpha for the test-retest was also excellent with a value of 0.86 (40).

In 2002, Gallagher et al. demonstrated the reliability of the MIST VR virtual reality simulator (48). They divided subjects into three groups of twelve; experienced laparoscopic surgeons, inexperienced laparoscopic surgeons and laparoscopic novices. Each subject completed ten trials on the MIST VR virtual reality simulator. They found that experienced laparoscopic surgeons performed the tasks significantly ($p<0.01$) faster, with less error, and more economy in the movement of the instruments and the use of diathermy, and with greater consistency in performance. The standardized coefficient alpha, a measure of the internal consistency of the MIST VR, ranged from $\alpha = 0.89$ to

0.98, showing a high internal consistency. Test-retest reliability was more varied with a range from r = 0.96 to r = 0.5. It was high for time, errors, and economy of diathermy, but low for economy of movement measures (48). Although virtual reality simulators have a degree of innate reliability, as there is no interrater or intrarater variability, test-retest reliability can and should be measured. To date, the reliability of the LapVR virtual reality simulator has not been proven.

When looking at the minimum requirements of validity and reliability for an effective assessment tool, this study and the literature have strongly demonstrated this for the video trainer and FLS. This is not yet the case with the virtual reality simulators, where there is minimal literature on the reliability of these machines. This is an area where more research is needed, especially with residency programs incorporating them more frequently into training curriculums.

iii. Flexibility

FLS has demonstrated that it can be used to evaluate clinical competency across a range of practice. It can be used to evaluate laparoscopic skills in both beginner residents and expert laparoscopic surgeons. It has been found to cover the domains considered necessary for assessing laparoscopic skill (8, 15, 16, 34-43). Our study supports this, as both construct and predictive validity were demonstrated for the video trainer and FLS. Additionally, the video trainer is very portable and the FLS tasks can easily be repeated many times in different locations without difficulty. Portability is important in residency training programs as device's that are easily portable can be made more available to trainees. It is even possible for trainees to have a video trainer box at home for practice purposes. This increases the amount of time resident's can spend practicing skills. Our study demonstrated only partial construct and predictive validity for the LapVR. Additionally the reliability of the LapVR has not been established. Thus the LapVR has not shown that it can be used to evaluate clinical competency across a range of practice. Although the tasks can be repeated very easily on the simulator, it is not as portable as the video trainer, and requires the availability of technical support in near proximity. The requirement of technical support staff and the lack of portability limit the availability of the virtual simulator for trainee practice.

iv. Comprehensiveness

The face and content validity of the manual skills component of FLS was assessed and confirmed by a group of 44 expert laparoscopic surgeons (15). Members of the SAGES FLS committee further evaluated the tasks for any skills considered fundamental to laparoscopic surgery that were not assessed with these five tasks. The only skills found to not be effectively covered were trocar placement and cannulation of a tubular structure (15). Initial trocar placement utilizes open surgical technique and does not rely on a specific laparoscopic skill set. A cannulation task could easily be added to the video trainer. Therefore the FLS manual skills tasks cover the essential manual skills required to competently perform laparoscopic surgery and provide a comprehensive assessment of skill. Our study supports this as both construct and predictive validity were demonstrated for the video trainer and FLS. The LapVR tasks have not yet shown that they cover the relevant objectives and assess the manual skills considered necessary to competently perform laparoscopic surgery. As we have demonstrated only partial construct and predictive validity of the LapVR we cannot say that it is comprehensive.

v. Feasibility

The video trainer and FLS are practical, easily portable and relatively inexpensive (37). The average set up for the video trainer and FLS practice materials is approximately $2000. In comparison, the virtual reality simulator is more difficult to transport, requires technical support, and costs on average $80,000 - $120,000 to purchase, as well as additional maintenance and software costs. Cost effectiveness is important to look at when evaluating an assessment tool. In order for an assessment to be cost effective, reliability, validity and educational impact need to be taken into account as well as monetary cost (99). As mentioned previously, these have been more thoroughly demonstrated for FLS and the video trainer than for the LapVR. In residency training programs where funds and resources are limited, video trainers are significantly more cost effective than virtual reality simulators.

Although not formally assessed, the trainees reported the video trainer much more acceptable to use than the virtual reality simulator. Many of the trainees found the virtual simulator tasks extremely unrealistic. For example, in the knot tying task the string floats through the air as if suspended in space and does not behave in a realistic manner. This is consistent with other formal (30) and informal (98) trainee perspectives in the literature. In order to improve the LapVR, the haptic feedback would need to be adjusted. With this in mind, the video trainer appears substantially more feasible than the virtual reality simulator.

vi. Timeliness

Both FLS and the virtual simulator assessments need to be performed as the tasks are being done. The virtual simulator automatically generates computer evaluations, although depending on the type of machine; only some of these parameters have been validated. An evaluator must go through the evaluation and extract the validated parameters. As they are stored on the computer hardware, it is possible to do this at a later time. For the FLS tasks, an evaluator is required to be present either in person or remotely (100) during the assessment. This presents the opportunity for immediate feedback to the trainee on their performance. The LapVR virtual reality simulator may offer more flexibility with the timing of the evaluation, as the evaluator does not have to be present during the assessment. However, extracting the validated parameters for the LapVR was much more labour intensive than assessing the FLS tasks. This could potentially delay feedback to the trainee if it were not done immediately following the assessment. Feedback to learners is more effective the closer in time it occurs to the interaction as it results in less performance recall degradation and assessment subjectivity (52, 53, 63). Therefore the delay in data extraction with the LapVR could be detrimental to the educational process. Software developers must improve the process of data extraction if the LapVR is to be an effective and timely instrument in surgical education.

vii. Accountability

The FLS scoring metrics meet the criteria for, and are currently used for high stakes examination. As of 2010, the American Board of Surgery requires its candidates to be FLS certified (34, 44). This assessment technique is defensible and accountable to all parties involved including the program and institution, students, educators, licensing authority and the community (8, 15, 16, 34-43). Our study supports this, as both construct and predictive validity were demonstrated for the FLS tasks on the video trainer. Currently there is not enough evidence for virtual reality to say that it is defensible and accountable to the above-mentioned stakeholders.

viii. Relevance

Both FLS and the LapVR offer the potential for either negative or positive assessments. For FLS if a trainee achieves a positive assessment result and demonstrates laparoscopic competency, they should be allowed to move onto performing laparoscopic procedures in the operating room. If the trainee achieves a negative result then a plan for remediation should be set in place before they move into the operating room. This could include a certain number of repetitions of the FLS tasks per day until a positive evaluation was obtained. This remediation plan could also be implemented for a resident who was performing poorly in the operating room. They could be sent back to the laboratory for simulator training until they were able to obtain an assessment demonstrating their laparoscopic competency. As validity and reliability have not consistently been demonstrated for the LapVR it should not currently be used as an assessment tool to advance or remediate trainees. If successful completion of these evaluations were mandatory prior to a trainee performing laparoscopic procedures in the operating room, the assessments would be viewed as an important part of the training process and would thus be a more effective assessment tool (53, 63). Additionally, in order for an assessment to be effective as an educational tool, the evaluation metrics need to provide meaningful feedback to the trainee. For FLS, it is possible for the evaluator to give feedback such as the trainee is too rough with the materials, cuts inaccurately or is too slow. The LapVR scoring metrics are not as uniformly educationally relevant. Some of the metrics provide meaningful feedback, for example overall time. Others such as non-dominant path length or non-dominant time are not as useful to the trainee. Re-examining these outcome metrics and coming up with more educationally meaningful ones would help improve the LapVR.

B. Why is the assessment being performed? What should be assessed?

The primary purpose of assessment with the video trainer and the virtual reality simulator is evaluation of skills objectives involving the psychomotor and technical skills that are required to competently perform laparoscopic surgery. Laparoscopic surgery provides a unique skill set that is particularly challenging for surgical trainees. In combination with reduced resident work hours and concern regarding the ethics of learning on patients, a method to assess a trainee's readiness to perform laparoscopic skills outside of the operating room is essential. Assessments provide feedback to the learner on their strengths and weaknesses, identifying areas requiring remediation or advancement potential. FLS manual skills performed on the video trainer meet the objectives deemed essential for competent laparoscopic skill (8, 15, 16, 34-43). Performance on these tasks can identify residents that are ready to progress in the operating room, as well as areas that may require further training and practice before being allowed to perform laparoscopic procedures on patients. The face and content validity of the LapVR manual skills have not been demonstrated or comprehensively assessed. The tasks are similar to the FLS tasks, however there were significant concerns regarding the realism of many of them. In order to determine if the LapVR is adequately assessing the psychomotor and technical skills required to competently perform laparoscopic surgery, a group of expert laparoscopists needs to evaluate the tasks for face and content validity. The LapVR does not currently meet the requirements of a valid and reliable assessment tool. It therefore does not meet the objectives essential for laparoscopic skill, and it cannot be recommended for assessment purposes at this time.

C. How will the assessment be categorized and judged?

The evaluation of the FLS tasks performed on the video trainer is categorized as a summative assessment. The performance on the tasks are used to decide whether or not the resident possess the necessary laparoscopic competence to advance to performing laparoscopic surgery in the operating room. The assessment is judged as criterion-referenced testing as the concern is that the trainee attains a minimal level of competence, not on their ranking within their peer group. FLS could also be used as a formative assessment to provide feedback to trainees in a less formal manner. Strengths and weaknesses could be identified, allowing the trainee to practice skills that were not as strong. The findings of our study support the use of FLS for both formative and summative assessments. Potentially the FLS skills could be used in a normative fashion to evaluate candidates applying for advanced laparoscopic fellowships. This would have to be interpreted cautiously since although it could demonstrate a candidates current laparoscopic technical ability, it would not be able to predict the candidates ability to acquire technical skill in the future, nor would it take into account a candidates overall clinical or operative judgment. At this point in time the LapVR has not demonstrated that is can be used for assessment of laparoscopic skills.

D. Who should perform the assessment?

Both the FLS tasks on the video trainer and the virtual reality simulator tasks could be evaluated by a non-physician who had been instructed in the various tasks and how the evaluations worked. For the video trainer tasks, the evaluator would need to be instructed in the various tasks, in how to explain them clearly to the trainees, in how the evaluation forms work, and in how to calculate the scoring metrics. This is currently what the formal FLS program does. For the LapVR virtual reality simulator, the computer automatically generates evaluations based on a number of parameters. The individual evaluating would need instructions regarding which validated parameters to extract and record. Alternatively this could be written into the software code. Once the evaluations were compiled, an attending surgeon could review them before deciding if a resident was ready for the operating room. An advantage to non-physicians performing these evaluations includes more flexibility as to when the assessments can occur. Physicians are extremely busy, and finding time to schedule these assessments is a major limiting factor. The ability to review the evaluations once they are compiled would expedite these assessments and help remove attending physician time as a major limiting factor. The potential also exists for residents to use these assessments for self-evaluation purposes. Once they have been correctly instructed in the use of the assessment tool, they would be able to perform self-evaluations. This would help trainees recognize strengths and weaknesses with their laparoscopic skills, and identify areas that needed work prior to entering the operating room. As this study only showed validity for the video trainer and FLS tasks, only the video trainer can be recommended for assessment purposes at this time.

E. When should the assessment occur?

Assessment of operative skills should occur continuously throughout a resident's training program, as it is a measure of skill development and evolvement. Trainees should be progressing and advancing their operative skills throughout their training program. It is important to continuously assess these skills to ensure trainees are progressing appropriately, and to identify and remediate appropriately individuals who are not. Assessment of laparoscopic skill should be performed prior to the resident performing laparoscopic surgery in the operating room. There is a unique skill set associated with laparoscopy, and it is important to ensure residents have mastered some of the basic laparoscopic skills before they participate in laparoscopic surgery in the operating room. This reduces risk to patients, ensures operating room efficiency, and decreases stress to the operating room team. The FLS manual skills performed on the video trainer meet the objectives deemed essential for competent laparoscopic skill and could be used to perform these assessments (8, 15, 16, 34-43). The findings of our study support this.

F. Project Significance

The reduction of residency duty hours and the increased cost of operating room time have greatly reduced the resources available to train residents to the same educational result as previously. This combined with the unique skill set afforded to laparoscopic surgery, increased public awareness and the ethics of teaching trainees basic skills on patients, has necessitated the development of cost-effective alternative methods of teaching and assessing residents laparoscopic skills outside of the operating theatre.

By demonstrating both strong construct and predictive validity for FLS, this study has contributed to the body of research validating FLS and the video trainer for assessment of laparoscopic skills. It has also shown that both the construct and predictive validity of the LapVR have been only partially demonstrated, and that more research and refinement would be required before implementing the LapVR as an assessment tool. With this evidence, and the fact that the FLS tasks and the video trainer are significantly less costly than the high cost LapVR, we can currently recommend FLS and the video trainer for assessment of surgical trainees laparoscopic skills over the LapVR virtual reality simulator. In order for the substantially higher cost LapVR to be recommended for surgical education purposes it would have to demonstrate that it provided a superior assessment of laparoscopic skill compared to FLS and the video trainer.

G. Project Limitations and Solutions

The present study has some limitations. First, the results are drawn from a single institution with 26 participants. Although this number is consistent with related educational literature and recommended guidelines, a multi-center study with a larger sample size would help strengthen the results. Second, there is no overall comprehensive LapVR score as in FLS. Therefore individual LapVR tasks must be compared to overall FLS score. The development of an overall LapVR score would allow a more direct comparison between the two forms of simulation. The company CAE Healthcare that developed the LapVR is currently investigating this. Third, there was some variability in the time between laparoscopic cholecystectomy and simulator tasks. Given the complicated nature of surgical residents schedules, opportunity sampling was used. Thus, setting up a specific time frame between the activities was not possible. There was a longer time interval for the junior than the senior residents, however many of the junior residents were rotating on non surgical off-service rotations, and it was felt that a longer interval between activities for the juniors would not be as detrimental as a significant period of time between tasks for the seniors. Additionally, the number of interim laparoscopic procedures was relatively low for both the senior and junior resident groups. Therefore, this should have had an insignificant impact on the results of this project.

H. Recommendations and Further Investigation

FLS is a well-validated and cost-effective method of assessing surgical trainees laparoscopic skills. It meets the criteria for an effective assessment tool, and offers flexibility in terms of when the assessment can occur and who can perform it. With the steep learning curves associated with learning to perform laparoscopic surgery, and the potential for devastating patient complications, surgical training programs should consider implementing the use of FLS manual skills for assessment prior to trainees performing laparoscopic procedures on humans. Development of an overall comprehensive score for the LapVR may help establish validity and reliability, and it would allow for a more direct comparison with overall FLS score. A significantly larger number of participants would be required for this, and a multi-center study may need to be considered.

Chapter 10 Conclusion

Construct validity was demonstrated for GOALS, OSATS, and overall FLS score on the video trainer. For the LapVR, it was a mixed picture. Construct validity for the LapVR was completely demonstrated for one of the five tasks, partially demonstrated for three of the five, and not demonstrated for one of the five. Compared to FLS, construct validity was much weaker and less thoroughly demonstrated for the LapVR.

Strong predictive validity was demonstrated for FLS and both GOALS, and OSATS scores. This is in contrast to the LapVR where predictive validity was difficult to determine. Strong predictive validity was only demonstrated for two of the five tasks and GOALS scores, and three of the five tasks and OSATS scores.

From this study, both construct and predictive validity were more thoroughly demonstrated for FLS and the video trainer than the LapVR virtual reality simulator. Overall FLS score explains more of the variation in operative score than the LapVR tasks, and it is currently validated for high stakes examination. There is currently not enough evidence to recommend utilizing the LapVR virtual reality simulator for assessment of laparoscopic skills. This would require further research and validation.

In residency training programs where funds and resources are often quite limited, the higher cost LapVR remains experimental in surgical education. Efforts instead should be focused on utilizing the well-validated, lower cost FLS and video trainer for assessment of laparoscopic skills.

References

1. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. American journal of surgery. 2010;199(1):115-20. Epub 2010/01/28.
2. Tedesco MM, Pak JJ, Harris EJ, Jr., Krummel TM, Dalman RL, Lee JT. Simulation-based endovascular skills assessment: the future of credentialing? Journal of vascular surgery. 2008;47(5):1008-1; discussion 14. Epub 2008/03/29.
3. Feldman LS, Sherman V, Fried GM. Using simulators to assess laparoscopic competence: ready for widespread use? Surgery. 2004;135(1):28-42. Epub 2003/12/25.
4. Krummel TM. Surgical simulation and virtual reality: the coming revolution. Annals of surgery. 1998;228(5):635-7. Epub 1998/12/02.
5. Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD. The effect of bench model fidelity on endourological skills: a randomized controlled study. The Journal of urology. 2002;167(3):1243-7. Epub 2002/02/08.
6. Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, et al. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. Annals of surgery. 2004;240(2):374-81. Epub 2004/07/27.
7. Gorman PJ, Meier AH, Rawn C, Krummel TM. The future of medical education is no longer blood and guts, it is bits and bytes. American journal of surgery. 2000;180(5):353-6. Epub 2001/01/04.
8. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. American journal of surgery. 1998;175(6):482-7. Epub 1998/06/30.
9. Fried GM. Lessons from the surgical experience with simulators: incorporation into training and utilization in determining competency. Gastrointestinal endoscopy clinics of North America. 2006;16(3):425-34. Epub 2006/08/01.
10. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. The British journal of surgery. 2004;91(12):1549-58. Epub 2004/11/18.
11. E.A.E.S. Guidelines: Training and assessment of competence. Surgical endoscopy. 1994;8:721-2.
12. SAGES. Granting of privileges for laparoscopic general surgery. American journal of surgery. 1991;161:324-5.
13. Philibert I, Friedmann P, Williams WT. New requirements for resident duty hours. JAMA : the journal of the American Medical Association. 2002;288(9):1112-4. Epub 2002/09/11.
14. Canadian Association of Interns and Residents. Canadian Patient and Physician Safety and Wellbeing: Resident Duty Hours. 2012.
15. Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G, et al. Proving the value of simulation in laparoscopic surgery. Annals of surgery. 2004;240(3):518-25; discussion 25-8. Epub 2004/08/21.

16. Edelman DA, Mattos MA, Bouwman DL. Impact of fundamentals of laparoscopic surgery training during medical school on performance by first year surgical residents. The Journal of surgical research. 2011;170(1):6-9. Epub 2011/06/24.
17. Munz Y, Kumar BD, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers: is one superior to the other? Surgical endoscopy. 2004;18(3):485-94. Epub 2004/01/31.
18. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Annals of surgery. 2002;236(4):458-63; discussion 63-4. Epub 2002/10/09.
19. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane database of systematic reviews (Online). 2009(1):CD006575. Epub 2009/01/23.
20. Zendejas B, Hernandez-Irizarry R, Farley DR. Does simulation training improve outcomes in laparoscopic procedures? Advances in surgery. 2012;46:61-71. Epub 2012/08/10.
21. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. The British journal of surgery. 2004;91(2):146-50. Epub 2004/02/05.
22. Fairhurst K, Strickland A, Maddern G. The LapSim virtual reality simulator: promising but not yet proven. Surgical endoscopy. 2011;25(2):343-55. Epub 2010/07/09.
23. Beyer L, Troyer JD, Mancini J, Bladou F, Berdah SV, Karsenty G. Impact of laparoscopy simulator training on the technical skills of future surgeons in the operating room: a prospective study. American journal of surgery. 2011;202(3):265-72. Epub 2011/06/11.
24. Torkington J, Smith SG, Rees BI, Darzi A. Skill transfer from virtual reality to a real laparoscopic task. Surgical endoscopy. 2001;15(10):1076-9. Epub 2001/12/01.
25. Kundhal PS, Grantcharov TP. Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. Surgical endoscopy. 2009;23(3):645-9. Epub 2008/07/16.
26. Loukas C, Nikiteas N, Schizas D, Lahanas V, Georgiou E. A head-to-head comparison between virtual reality and physical reality simulation training for basic skills acquisition. Surgical endoscopy. 2012;26(9):2550-8. Epub 2012/04/06.
27. Hyltander A, Liljegren E, Rhodin PH, Lonroth H. The transfer of basic skills learned in a laparoscopic simulator to the operating room. Surgical endoscopy. 2002;16(9):1324-8. Epub 2002/05/04.
28. Ghaderi I, Vaillancourt M, Sroka G, Kaneva PA, Seagull FJ, George I, et al. Performance of simulated laparoscopic incisional hernia repair correlates with operating room performance. American journal of surgery. 2011;201(1):40-5. Epub 2010/12/21.
29. Andreatta PB, Woodrum DT, Birkmeyer JD, Yellamanchilli RK, Doherty GM, Gauger PG, et al. Laparoscopic skills are improved with LapMentor training: results of a randomized, double-blinded study. Annals of surgery. 2006;243(6):854-60; discussion 60-3. Epub 2006/06/15.
30. Hamilton EC, Scott DJ, Fleming JB, Rege RV, Laycock R, Bergen PC, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. Surgical endoscopy. 2002;16(3):406-11. Epub 2002/04/03.

31. Scott DJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Valentine RJ, et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? Journal of the American College of Surgeons. 2000;191(3):272-83. Epub 2000/09/16.
32. Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, Grantcharov T, et al. Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. Annals of surgery. 2011;253(6):1216-22. Epub 2011/04/26.
33. Palter VN, Orzech N, Reznick RK, Grantcharov TP. Validation of a structured training and assessment curriculum for technical skill acquisition in minimally invasive surgery: a randomized controlled trial. Annals of surgery. 2013;257(2):224-30. Epub 2012/09/28.
34. Vassiliou MC, Feldman LS. Objective assessment, selection, and certification in surgery. Surgical oncology. 2011;20(3):140-5. Epub 2010/11/13.
35. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery. 2004;135(1):21-7. Epub 2003/12/25.
36. Soper NJ, Fried GM. The fundamentals of laparoscopic surgery: its time has come. Bulletin of the American College of Surgeons. 2008;93(9):30-2. Epub 2009/05/28.
37. Fried GM. FLS assessment of competency using simulated laparoscopic tasks. Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2008;12(2):210-2. Epub 2007/10/24.
38. Scott DJ, Ritter EM, Tesfay ST, Pimentel EA, Nagji A, Fried GM. Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training. Surgical endoscopy. 2008;22(8):1887-93. Epub 2008/02/14.
39. McCluney AL, Vassiliou MC, Kaneva PA, Cao J, Stanbridge DD, Feldman LS, et al. FLS simulator performance predicts intraoperative laparoscopic skill. Surgical endoscopy. 2007;21(11):1991-5. Epub 2007/06/27.
40. Vassiliou MC, Ghitulescu GA, Feldman LS, Stanbridge D, Leffondre K, Sigman HH, et al. The MISTELS program to measure technical skill in laparoscopic surgery : evidence for reliability. Surgical endoscopy. 2006;20(5):744-7. Epub 2006/03/02.
41. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM. Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. Surgical endoscopy. 2003;17(6):964-7. Epub 2003/03/27.
42. Keyser EJ, Derossis AM, Antoniuk M, Sigman HH, Fried GM. A simplified simulator for the training and evaluation of laparoscopic skills. Surgical endoscopy. 2000;14(2):149-53. Epub 2000/02/05.
43. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM. Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. Journal of the American College of Surgeons. 2004;198(1):105-10. Epub 2003/12/31.
44. Edelman DA, Mattos MA, Bouwman DL. FLS skill retention (learning) in first year surgery residents. The Journal of surgical research. 2010;163(1):24-8. Epub 2010/07/08.
45. Gallagher AG, Lederman AB, McGlade K, Satava RM, Smith CD. Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using

criteria levels based on expert performance. Surgical endoscopy. 2004;18(4):660-5. Epub 2004/03/18.
46. Langelotz C, Kilian M, Paul C, Schwenk W. LapSim virtual reality laparoscopic simulator reflects clinical experience in German surgeons. Langenbeck's archives of surgery / Deutsche Gesellschaft fur Chirurgie. 2005;390(6):534-7. Epub 2005/07/30.
47. Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World journal of surgery. 2001;25(11):1478-83. Epub 2002/01/05.
48. Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surgical endoscopy. 2002;16(12):1746-52. Epub 2002/07/26.
49. Grantcharov TP, Rosenberg J, Pahle E, Funch-Jensen P. Virtual reality computer simulation. Surgical endoscopy. 2001;15(3):242-4. Epub 2001/05/19.
50. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Annals of surgery. 2007;245(6):992-9. Epub 2007/05/25.
51. van Dongen KW, Tournoij E, van der Zee DC, Schijven MP, Broeders IA. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? Surgical endoscopy. 2007;21(8):1413-7. Epub 2007/02/13.
52. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: From selection to certification. Current problems in surgery. 2002;39(6):573-659. Epub 2002/05/31.
53. Turnbull J, Gray J, MacFadyen J. Improving in-training evaluation programs. Journal of general internal medicine. 1998;13(5):317-23. Epub 1998/06/05.
54. Harden R. How to Assess Students: An Overview. Medical Teacher. 1979;1(2):65-70.
55. Reznick RK. Teaching and testing technical skills. American journal of surgery. 1993;165(3):358-61. Epub 1993/03/01.
56. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surgical endoscopy. 2003;17(10):1525-9. Epub 2003/09/23.
57. Education ACoSDo. 2011; Available from: http://www.facs.org/education/technicalskills/faqs/faqs.html.
58. Reber A. The Penguin dictionary of psychology. 2nd ed. New York: Penguin; 1995.
59. van Empel PJ, van der Veer WM, van Rijssen LB, Cuesta MA, Scheele F, Bonjer HJ, et al. Mapping the maze of minimally invasive surgery simulators. Journal of laparoendoscopic & advanced surgical techniques Part A. 2012;22(1):51-60. Epub 2011/12/08.
60. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. The British journal of surgery. 2010;97(7):972-87. Epub 2010/07/16.
61. Benton T. Exploring equivalent forms reliability using a key stage 2 reading test. Research Papers in Education. 2012;28(1):57-74.
62. Philips SE. Legal issues in performance assessment. West's Educ Law Q. 1993;2(2):329-58.

63. Hunt DD. Functional and dysfunctional characteristics of the prevailing model of clinical evaluation systems in North American medical schools. Academic medicine : journal of the Association of American Medical Colleges. 1992;67(4):254-9. Epub 1992/04/01.
64. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Academic medicine : journal of the Association of American Medical Colleges. 1998;73(9):993-7. Epub 1998/10/06.
65. Falchikov B, Boud D. Student self-assessment in higher education: a meta analysis. Review of Educational Research. 1989;70:395-430.
66. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Academic medicine : journal of the Association of American Medical Colleges. 1991;66(12):762-9. Epub 1991/12/01.
67. Mandel LS, Goff BA, Lentz GM. Self-assessment of resident surgical skills: is it feasible? American journal of obstetrics and gynecology. 2005;193(5):1817-22. Epub 2005/11/02.
68. Speyer R, Pilz W, Van Der Kruis J, Brunings JW. Reliability and validity of student peer assessment in medical education: a systematic review. Med Teach. 2011;33(11):e572-85. Epub 2011/10/26.
69. Risucci DA, Tortolani AJ, Ward RJ. Ratings of surgical residents by self, supervisors and peers. Surgery, gynecology & obstetrics. 1989;169(6):519-26. Epub 1989/12/01.
70. Secomb J. A systematic review of peer teaching and learning in clinical education. Journal of clinical nursing. 2008;17(6):703-16. Epub 2007/12/01.
71. Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. Teach Learn Med. 1990;2:58-76.
72. Woolliscroft JO, Howell JD, Patel BP, Swanson DB. Resident-patient interactions: the humanistic qualities of internal medicine residents assessed by patients, attending physicians, program supervisors, and nurses. Academic medicine : journal of the Association of American Medical Colleges. 1994;69(3):216-24. Epub 1994/03/01.
73. Butterfield PS, Mazzaferri EL, Sachs LA. Nurses as evaluators of the humanistic behavior of internal medicine residents. Journal of medical education. 1987;62(10):842-9. Epub 1987/10/01.
74. Fitts PM, Posner MI. Human Performance. Belmont, CA: Brooks/Cole; 1967.
75. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. The New England journal of medicine. 2006;355(25):2664-9. Epub 2006/12/22.
76. Fried GM, Feldman LS. Objective assessment of technical performance. World journal of surgery. 2008;32(2):156-60. Epub 2007/06/15.
77. Gray JD. Global rating scales in residency education. Academic medicine : journal of the Association of American Medical Colleges. 1996;71(1 Suppl):S55-63. Epub 1996/01/01.
78. Beard JD. Assessment of surgical skills of trainees in the UK. Annals of the Royal College of Surgeons of England. 2008;90(4):282-5. Epub 2008/05/22.
79. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. American journal of surgery. 2005;190(1):107-13. Epub 2005/06/24.

80. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. BJOG : an international journal of obstetrics and gynaecology. 2008;115(7):908-16. Epub 2008/05/20.
81. Scott DJ, Rege RV, Bergen PC, Guo WA, Laycock R, Tesfay ST, et al. Measuring operative performance after laparoscopic skills training: edited videotape versus direct observation. Journal of laparoendoscopic & advanced surgical techniques Part A. 2000;10(4):183-90. Epub 2000/09/21.
82. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. The British journal of surgery. 1997;84(2):273-8. Epub 1997/02/01.
83. MacRae H, Regehr G, Leadbetter W, Reznick RK. A comprehensive examination for senior surgical residents. American journal of surgery. 2000;179(3):190-3. Epub 2000/05/29.
84. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. American journal of surgery. 1997;173(3):226-30. Epub 1997/03/01.
85. Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. Journal of the American College of Surgeons. 2007;204(2):308-13. Epub 2007/01/27.
86. Darzi A, Mackay S. Assessment of surgical competence. Quality in health care : QHC. 2001;10 Suppl 2:ii64-9. Epub 2001/11/09.
87. Hammond I, Karthigasu K. Training, assessment and competency in gynaecologic surgery. Best practice & research Clinical obstetrics & gynaecology. 2006;20(1):173-87. Epub 2005/11/10.
88. Undre S, Darzi A. Laparoscopy simulators. Journal of endourology / Endourological Society. 2007;21(3):274-9. Epub 2007/04/21.
89. Rassweiler J, Klein J, Teber D, Schulze M, Frede T. Mechanical simulators for training for laparoscopic surgery in urology. Journal of endourology / Endourological Society. 2007;21(3):252-62. Epub 2007/04/21.
90. van der Meijden OA, Schijven MP. The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review. Surgical endoscopy. 2009;23(6):1180-90. Epub 2009/01/02.
91. Strom P, Hedman L, Sarna L, Kjellin A, Wredmark T, Fellander-Tsai L. Early exposure to haptic feedback enhances performance in surgical simulator training: a prospective randomized crossover study in surgical residents. Surgical endoscopy. 2006;20(9):1383-8. Epub 2006/07/11.
92. Tholey G, Desai JP, Castellanos AE. Force feedback plays a significant role in minimally invasive surgery: results and analysis. Annals of surgery. 2005;241(1):102-9. Epub 2004/12/29.
93. Lamata P, Gomez EJ, Sanchez-Margallo FM, Lamata F, del Pozo F, Uson J. Tissue consistency perception in laparoscopy to define the level of fidelity in virtual reality simulation. Surgical endoscopy. 2006;20(9):1368-75. Epub 2006/07/22.
94. Savoldelli GL, Naik VN, Joo HS, Houston PL, Graham M, Yee B, et al. Evaluation of patient simulator performance as an adjunct to the oral examination for senior anesthesia residents. Anesthesiology. 2006;104(3):475-81. Epub 2006/03/02.

95. Vergis A, Gillman L, Minor S, Taylor M, Park J. Structured assessment format for evaluating operative reports in general surgery. American journal of surgery. 2008;195(1):24-9. Epub 2007/12/18.
96. Gillman LM, Vergis A, Park J, Minor S, Taylor M. Structured operative reporting: a randomized trial using dictation templates to improve operative reporting. American journal of surgery. 2010;199(6):846-50. Epub 2010/07/09.
97. Rivard J. The Effects of Blocked Versus Random Task Practice Schedules on the Acquisition, Retention, and Transfer of Surgical Skills. FGS - Electronic Theses & Dissertations (Public): University of Manitoba; 2013.
98. Iwata N, Fujiwara M, Kodera Y, Tanaka C, Ohashi N, Nakayama G, et al. Construct validity of the LapVR virtual-reality surgical simulator. Surgical endoscopy. 2011;25(2):423-8. Epub 2010/06/30.
99. Schuwirth LVvd, C. Cost effectiveness in medical education. Google Books: Radcliffe; 2010.
100. Okrainec A, Henao O, Azzie G. Telesimulation: an effective method for teaching the fundamentals of laparoscopic surgery in resource-restricted countries. Surgical endoscopy. 2010;24(2):417-22. Epub 2009/07/01.

Appendix 1. Objective Structured Assessment of Technical Skills (OSATS) Assessment Form for Laparoscopic Cholecystectomy

Date:________________________ Operator:________________________

Attending:________________________ Level of Training:________________________

Objective Structured Assessment of Technical Skills (OSATS)

Performance characteristic	Rating					
Respect for tissue	1 Frequently used unnecessary force on tissue or caused damage	2	3 Careful handling of tissue but occasionally caused inadvertent damage	4	5 Consistently handled tissues appropriately with minimal damage	NA
Time and motion	1 Many unnecessary moves	2	3 Efficient time/motion but some unnecessary moves	4	5 Clear economy of motion and maximum efficacy	NA
Instrument handling	1 Repeatedly makes tentative awkward or inappropriate moves with instruments	2	3 Competent use of instruments but occasionally stiff or awkward	4	5 Fluid moves with instruments and no awkwardness	NA
Plane of dissection	1 Frequently entered or followed incorrect dissection plan	2	3 Correct plane of dissection most of the time but occasionally went off plane	4	5 Entered and followed correct tissue plane	NA
Knowledge of instruments	1 Frequently asked for wrong instrument or used inappropriate instrument	2	3 Knew names of most instruments and used appropriate tool for task	4	5 Obviously familiar with the instruments and their names	NA
Flow of procedure	1 Frequently stopped procedure and seemed unsure of next move	2	3 Demonstrated some forward planning with reasonable progression of procedure	4	5 Planned course of procedure effortless from one move to the next	NA
Use of assistants	1 Failed to use assistants	2	3 Appropriate use of assistants most of the time	4	5 Strategically used assistants to the best advantage at all times	NA
Knowledge of procedure	1 Required specific instruction at most steps	2	3 Knew all important steps of the procedure	4	5 Familiar with all aspects of the procedure	NA
Overall performance	1 Unable to perform procedure independently	2	3 Competent, could perform procedure with minimal teaching assistance	4	5 Clearly superior, able to perform procedure independently with confidence	NA
Overall: Pass / Fail	Fail				Pass	
Total Score						**/45**

*Modified from Reznick R, Regehr G, Macrae H, et al. Testing technical skill via an innovative bench station examination. Am J Surg 1996; 173:226-230. 1=worst possible score, 5=best possible score

Appendix 2. Global Operative Assessment of Laparoscopic Skills (GOALS) Assessment Form for Laparoscopic Cholecystectomy

Date: ____________________ Operator: ____________________
Attending: ____________________ Level of Training: ____________________

Global Operative Assessment of Laparoscopic Skills – GOALS

GLOBAL RATING SCALE - **GRS**

1. Depth Perception Score:

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

2. Bimanual Dexterity Score:

1. Uses only one hand, ignores non-dominant hand, poor coordination between hands
2.
3. Uses both hands, but does not optimize interaction between hands
4.
5. Expertly utilizes both hands in a complimentary manner to provide optimal exposure

3. Efficiency Score:

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

4. Tissue Handling Score:

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

5. Autonomy Score:

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

Total **/25**

VISUAL ANALOGUE SCALE - **VAS**

Degree of difficulty. Place an "X" along the line:

Extremely easy. Planes well-defined, no scar tissue/edema. |————————————| Extremely difficult. Invisible planes and excessive scarring

Overall competence (dissection of the gallbladder from the liver bed). The operator:

Was unable to complete the task with maximum guidance |————————————| Could perform the task safely and independently (fully competent)

Completed by: ____________________

Appendix 3. Form Used to Obtain Informed Consent

RESEARCH PARTICIPANT INFORMATION AND CONSENT FORM

Title of Study: Do Fundamentals of Laparoscopic Surgery (FLS) and LapSim evaluation metrics predict intra-operative performance?

Principle Investigator
Dr. Ashley Vergis, St. Boniface General Hospital Z3039-409 Tache Avenue Winnipeg Manitoba R2H2A6 237-2568

Co-Investigator
Dr. Jason Park, St. Boniface General Hospital Z3031-409 Tache Avenue Winnipeg Manitoba R2H2A6 235-3255

You are being asked to participate in a research study. Please take your time to review this consent form and discuss any questions you may have with the study staff. You may take your time to make your decision about participating in this study and you may discuss it with your friends, family or (if applicable) your doctor before you make your decision. This consent form may contain words that you do not understand. Please ask the study staff to explain any words or information that you do not clearly understand.

Purpose of Study
This research is being conducted to study if Fundamentals of Laparoscopic Surgery (FLS) and LapSim ability correlate with resident intra-operative performance using a live cholecystectomy model. A total of 30 participants will participate in this study.

Study Procedures
All Post Graduate Year (PGY) 1 General surgery and Urology residents at the University of Manitoba currently complete a 2-day basic laparoscopic skills course as part of their core curriculum. The session includes exposure to a limited number of box trainers and performance of a laparoscopic cholecystectomy using a live porcine model. The course will be expanded to 3 days and short instructional videos demonstrating correct task performance will be shown to familiarize residents with the procedures and equipment at the beginning of each session. Residents will be assigned a study number and recorded each performing: the 5 FLS tasks (peg transfer, pattern cutting, ligating loop, intracorporeal and extracorporeal suturing) on day 1, a simulated laparoscopic cholecystectomy and basic skills tasks (ranging from camera navigation to laparoscopic suturing) on LapSim (day 2), and an in vivo laparoscopic cholecystectomy using a live porcine model on day 3.

FLS task performance will be evaluated using standardized FLS metrics by a research assistant. Validated LapSim metrics including time, error score (tissue damage and millimeters of tissue damage), and economy of movement parameters (path length and angular path) will be measured by the simulator. Finally, two blinded staff evaluators will then watch and score each resident porcine video using the previously validated Objective Structured Assessment of Technical Skills (OSATS) ratings scale or Global Operative Assessment of Laparoscopic Skills (GOALS) ratings scale.

Participation in this study will be for approximately 3 days.

The researcher may decide to take you off this study if funding is stopped. If this is the case, the regular 2 day course which is part of the core curriculum will continue.

PGY 2 -5 general surgery residents will be evaluated each performing: the 5 FLS tasks (peg transfer, pattern cutting, ligating loop, intracorporeal and extracorporeal suturing), a simulated laparoscopic cholecystectomy and basic skills tasks (ranging from camera navigation to laparoscopic suturing) on LapSim, and an in vivo laparoscopic cholecystectomy using a live human model. All general surgery residents are currently evaluated using either the previously validated Objective Structured Assessment of Technical Skills (OSATS) ratings scale or Global Operative Assessment of Laparoscopic Skills (GOALS) ratings scale when they perform a laparoscopic cholecystectomy on a human. One of these evaluations for each resident will now be included for study purposes to measure intraoperative performance.

You can stop participating at any time. However, if you decide to stop participating in the study, we encourage you to talk to the study staff first.

Results will be provided to participants after evaluators have completed the scoring process. Results will be mailed to each participant. It is anticipated that results will be available approximately six months after the study is completed.

Risks and Discomforts

There are no anticipated risks or discomforts associated with this study. Residents complete a 2-day version of this course as part of the PGY1 core curriculum in General Surgery and Urology. They are evaluated as part of the residency program. This course is aimed at helping to better train residents in the fundamentals of laparoscopic surgery, an integral part of both the General Surgery and Urology residency training programs. There are no anticipated risks or discomforts associated with PGY2-5 general surgery residents completing this study. There are no risks associated with completing the FLS or LapSim tasks, and the laparoscopic cholecystectomy data is already collected as part of the residency training curricula.

Benefits
There may or may not be a direct benefit to you from participating in this study. We hope the information learned from this study will benefit other people with laparoscopic training methods and programs used to teach residents in the future.

Costs
All the procedures, which will be performed as part of this study, are provided at no cost to you.

Payment for participation
You will receive no payment or reimbursement for any expenses related to taking part in this study.

Alternatives
Not applicable.

Confidentiality
Information gathered in this research study may be published or presented in public forums, however your name and other identifying information will not be used or revealed. All study related documents will bear only your assigned study number. Despite efforts to keep your personal information confidential, absolute confidentiality cannot be guaranteed. Your personal information may be disclosed if required by law.

Participant videos and all electronic data will be stored on a study computer, which will be encrypted and password protected. Only the researchers involved with the study will have access to this information. Participant videos and data will be kept for seven years. The computer will be kept in a locked and secure office.

The University of Manitoba Health Research Ethics Board may review records related to the study for quality assurance purposes.

All records will be kept in a locked secure area and only those persons identified will have access to these records. If any of your research records need to be copied to any of the above, your name and all identifying information will be removed. No information revealing any personal information such as your name, address or telephone number will leave the researcher's office at the St. Boniface General Hospital.

Voluntary Participation/Withdrawal from the Study
Your decision to take part in this study is voluntary. If the study staff feel that it is in your best interest to withdraw you from the study, they will remove you without your consent. Your performance evaluation will not be affected by your decision to participate. If you decide not to participate in the study, you will complete the course as per the core curriculum for PGY1 residents in General Surgery and Urology, however you will not be video taped and evaluated for research/study purposes. None of your data will be saved and used in the study. There will be no penalty for residents who do not wish to

participate in this study.

We will tell you about any new information that may affect your health, welfare, or willingness to stay in this study.

Medical Care for Injury Related to the Study
There are no anticipated risks of injury or illness related to this study.

Questions
You are free to ask any questions that you may have about your treatment and your rights as a research participant. If any questions come up during or after the study or if you have a research-related injury, contact the study doctor and the study staff: Dr. Sarah Steigerwald 936-0944 (pager)

For questions about your rights as a research participant, you may contact The University of Manitoba, Bannatyne Campus Research Ethics Board Office at (204) 789-3389

Do not sign this consent form unless you have had a chance to ask questions and have received satisfactory answers to all of your questions.

Statement of Consent
I have read this consent form. I have had the opportunity to discuss this research study with Dr. Ashley Vergis and or his/her study staff. I have had my questions answered by them in language I understand. The risks and benefits have been explained to me. I believe that I have not been unduly influenced by any study team member to participate in the research study by any statements or implied statements. Any relationship (such as employer, supervisor or family member) I may have with the study team has not affected my decision to participate. I understand that I will be given a copy of this consent form after signing it. I understand that my participation in this study is voluntary and that I may choose to withdraw at any time. I freely agree to participate in this research study.

I understand that information regarding my personal identity will be kept confidential, but that confidentiality is not guaranteed. I authorize the inspection of any of my records that relate to this study by The University of Manitoba Research Ethics Board, for quality assurance purposes.

By signing this consent form, I have not waived any of the legal rights that I have as a participant in a research study.

I agree to be contacted for future follow-up in relation to this study,
Yes _ No _

Participant signature__________________________ Date___________ (day/month/year)
Participant printed name: _____________________________

Appendix 4. Questionnaire Given to Study Participants

Participant Questionnaire

Identification number: ______________________________________

Date: _______________________

1. Residency year (1 – 5) ____

2. Age ____

3. Gender F / M

4. Dominant hand L / R

5. How much experience do you have with laparoscopic procedures?

   Number of laparoscopic procedures observed ____

   Number of procedures acting as camera-person ____

   I have performed operative procedures Y / N

   If yes, please describe number and nature of performance.

6. How much laparoscopic experience have you had since you performed the first part of this study (either the FLS/LapVR or the laparoscopic cholecystectomy)?

7. Have you performed any parts of the Fundamental of Laparoscopic Surgery (FLS) program before? Y / N

   If yes, please describe nature of performance.

8. How much experience since you performed the first part of this study (either the FLS/LapVR or the laparoscopic cholecystectomy)?

9. Have you practiced on any other laparoscopic box trainers or computer-based simulators before? Y / N

If yes, please describe nature of performance.

10. How much experience since you performed the first part of this study (either the FLS/LapVR or the laparoscopic cholecystectomy)?