FLS Simulator Training to Proficiency Improves Laparoscopic Performance in the Operating Room – a Randomized Controlled Trial

Gideon Sroka MD

Division of Surgical Research

McGill University, Montreal

June 2009

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Experimental Surgery

Table of Contents

Abstract... 3

Résumé... 4

A. Introduction.. 5

B. Literature review... 11

I – Objective Assessment of Technical Performance 12

1. Assessment of performance in simulators.... 13

2. Assessment of operative performance......... 16

II- Skill Transfer from Simulator to Operating Room 25

III- Simulation training in basic laparoscopic skills.. 35

C. FLS Simulator Training to Proficiency Improves Laparoscopic Performance in the Operating Room – a Randomized Controlled Trial................................ 40

I- Contribution of Authors.............................. 41

II- Introduction... 42

III- Methods.. 44

IV- Results... 49

V- Discussion.. 51

D. Conclusion.. 58

E. Tables and Figures.. 63

F. References.. 72

Acknowledgements... 82

Appendix I- FLS simulator and tasks.............................. 83

Appendix II- FLS performance metrics............................ 85

Appendix III - GOALS instrument................................... 86

Abstract

There is growing interest in the use of simulation for surgical skills training and evaluation. The purpose of this study was to assess whether training to proficiency with the FLS laparoscopic simulator would result in improved performance in the operating room (OR). GOALS, a validated tool, was used to measure clinical operating room performance. Nineteen junior residents underwent baseline FLS-testing and GOALS evaluation during elective laparoscopic cholecystectomy. Those with GOALS scores≤15 were randomly assigned to training (n=9) or control (n=8) groups. An FLS proficiency-based curriculum was used in the training group. Scoring on FLS and in the OR was repeated at the end of the study period. Evaluators were blinded to randomization status. Sixteen residents completed the study. There were no differences in baseline simulator or OR scores. After training, simulator scores were higher in the training compared to control group. At the final assessment, the training group improved their OR performance significantly more than the control. The observed improvement was from novice to intermediate level of residency. These results show the transferability of basic laparoscopic skills gained on a physical simulator to the OR and emphasize the value of laparoscopic simulators for training purposes.

Résumé

Il y a un intérêt grandissant pour l'utilisation de la simulation à des fins de formation et d'évaluation des compétences de chirurgie. Le but de cette étude consistait à déterminer si une formation menant jusqu'à la compétence, effectuée sur le simulateur FLS, produirait un meilleur rendement en salle d'opération. Le rendement clinique a été mesuré à l'aide de GOALS, un outil validé. Initialement, dix-neuf résidents juniors ont subi un test FLS ainsi qu'une évaluation GOALS effectuée en salle d'opération pendant l'exécution d'une cholécystectomie par laparoscopie non urgente. Les résidents ayant un score GOALS ≤ 15 ont été répartis au hasard entre le groupe de formation (n=9) et le groupe témoin (n=8). Le groupe de formation a suivi un programme centré sur la compétence en matière de FLS. Les scores FLS et les scores en salle d'opération on été évalués de nouveau à la fin de la période d'étude. Les évaluations ont été effectuées en aveugle. Seize résidents ont terminé l'étude. Aux tests initiaux, aucune différence n'a été constatée entre les scores FLS et les scores en salle d'opération. Après la formation, les scores FLS étaient plus élevés chez le groupe de formation comparativement au groupe témoin. À l'évaluation finale, le groupe de formation avait amélioré sa performance en salle d'opération de façon significative comparativement au groupe témoin. L'amélioration observée s'est traduite par un passage du niveau de résidence novice au niveau intermédiaire. Ces résultats démontrent que les compétences de base en laparoscopie acquises à l'aide d'un simulateur peuvent être transférées en salle d'opération. De plus, ils soulignent la valeur des simulateurs de laparoscopie en tant qu'outils de formation.

SECTION A - INTRODUCTION

On September 12 1985, Dr Erich Muhe of Boblingen, Germany, performed the first laparoscopic cholecystectomy. His achievement was not acknowledged by the German Surgical Society until 7 years later, when he received their highest award. In 1987, Philip Mouret from France and 4 other European surgeons performed the operation independently and made it public for the first time [1]. The laparoscopic revolution in general surgery had begun. Practicing surgeons rushed to acquire the new technique in weekend courses and incorporated it into their practices. Some of these courses were good, and included instruction with inanimate or animal models, but proficiency could not have been gained in such a short time [2].

It was very soon clear that a different set of surgical skills was required for laparoscopic surgery: the struggle with depth perception while interpreting the two dimensional image on the monitor, and the control of the long instruments that move to the opposite direction from their controlling hand, were among the completely new challenges for the general surgeon [3]. The gap between the pressure to adopt the technique, originating from industry, as well as from patients, coupled with the inadequate educational opportunities available for practicing surgeons, was associated with a high incidence of complications during adoption of laparoscopic cholecystectomy, most notably bile duct injuries [4]. In response to this situation, the Society of American Gastrointestinal Endoscopic Surgeons

(SAGES) developed guidelines for credentialing surgeons in laparoscopic cholecystectomy which required training in animal models, and proctoring by an experienced surgeon during the first cases [5]. The American College of Surgeons (ACS) reorganized its Division of Education and in 1994 added the Committee on Emerging Surgical Technology and Education, with patient safety as a main concern with regard to the introduction of new surgical technologies. As more surgeons began to apply laparoscopic techniques for more complex abdominal operations (e.g. carcinoma of the colon), it became clear that safe introduction of innovations requiring the acquisition of new skills would require a change in the traditional surgical training paradigm of "see one, do one, teach one".

By the mid- 1990's, in addition to safety concerns made prominent by the introduction of laparoscopy, other pressures in surgical training emerged. With shortening of resident working hours and the increasing cost of (or decreased access to) OR time, there was increasing interest in an approach that would enable laparoscopic skills training outside of the OR. The use of simulation for surgical skills training in general was in rapid evolution at the time. A unique collaborative group of surgeons and computer scientists from France presented their work on the application of virtual reality simulation in hepatic surgery[6], and gave the surgical community a view to the future. Simulation had already been used in medical education in a variety of settings. Paramedical personnel were taught triage and assessment skills using this technique. Simulated scenarios were built into Advanced Trauma Life Support (ATLS) and Advanced Cardiac Life Support (ACLS) courses, to teach and test skills[7]. Mannequin-based simulations were being used in anesthesia training to ensure clinical exposure to unusual situations, and also in team training in trauma resuscitations and in teaching crisis management skills[8].

In 1998, Derossis et al introduced the MISTELS (McGill Inanimate System for Training and Evaluation of Laparoscopic Skills), a simple laparoscopic trainer box connected to a video monitor. Seven tasks simulating a variety of fundamental laparoscopic skills were created with performance metrics and evidence for "known-groups" construct validity of the metrics was reported. The metrics rewarded efficiency, but also penalized errors. Simulator performance correlated strongly with in vivo performance in a porcine model, in which similar tasks and evaluation metrics were used[9]. Further validation of MISTELS continued in different institutions and countries, after which two tasks were removed, leaving five tasks: Transferring, Pattern cutting, Placement and securing of a ligating loop, and placement of an interrupted suture with both extracorporeal and intracorporeal knot tying[10]. In 2004, MISTELS was incorporated by SAGES and the ACS as the manual skills component of a new comprehensive program for education and assessment of the basic Fundamentals of Laparoscopic Surgery (FLS)[11]. Subsequently, it has been referred to as the FLS simulator (Appendix I). Of note, FLS certification is now mandated by the American Board of Surgery[12] and the Royal Australasian College of Surgeons[13].

While the reliability and validity of the simulator performance metrics was demonstrated (Appendix II), the relationship between simulator and OR performance ("predictive validity") had not been demonstrated. In order to measure this relationship, a method to objectively assess surgical performance in the operating room was required. In 1997, the surgical education research group from the University of Toronto developed OSATS – Objective Structured Assessment of Technical Skills- an eight-station bench examination of technical skills in residents. Residents were assessed at each station using two methods: a task-specific checklist and global rating scales. The Global Rating Scale (GRS), which is now widely referred to as “OSATS”, correlated very well with multi-detailed checklists for the various tasks, but had higher reliability and was simpler to use[14]. Although OSATS has become the most widely used means to assess technical performance in the OR, it was developed and validated for open surgery. In 2004, Vassiliou et al from McGill University developed GOALS – Global Operative Assessment of Laparoscopic Skills - a laparoscopy-specific GRS to measure technical performance in the OR[15]. GOALS consists of 5 domains each evaluated on a 5 point Likert scale, with anchor descriptors for scores 1,3 and 5 for each domain (appendix III). GOALS was found to have strong construct validity and inter-rater reliability, and has an advantage over checklists or Visual Analogue Scales (VAS) in the sense that it provides residents with formative and summative feedback. It also allowed the researchers to demonstrate the predictive validity of the FLS simulator[16]. However, a missing link remained: does training on the FLS simulator produce better laparoscopic performance in real human operations? The objective of the study reported in this thesis is to assess whether training to a proficiency level on the FLS simulator improves junior surgical residents' operating room laparoscopic performance. . In order to situate the relevance of the study and its design, previous literature concerning surgical performance assessment and the transfer of skills from laparoscopic simulators to the operating room will first be reviewed. A manuscript reporting the study will then be presented; this was presented at the Association for Surgical Education in Salt Lake City, UT, in April 2009, and has been accepted for publication in the American Journal of Surgery.

SECTION B- LITERATURE REVIEW

I - Objective assessment of surgical performance

On July 2001, the Metrics for Objective Assessment of Surgical Skills Workshop held an international assembly of experts in the field joined by representatives of official bodies involved in surgical education, evaluation and certification. They acknowledged the large pool of scientific research on capabilities for safety in aviation, transportation, nuclear power and other fields, and suggested that those should be investigated for medical applications[17]. They emphasized the advantages of simulators for training, with their ability to automatically and objectively assess performance, providing real-time feedback and continuous tracking of improvement (learning curve). They made it clear that the type of simulator is less important than its associated curriculum, but noted that there is not yet a single, coherent “core curriculum” that is agreed upon. More importantly, they highlighted the large variability in the outcome measures for performance analysis. his section reviews the literature concerning measurement of surgical performance. Some approaches are used only in simulators, whereas others can be used both for simulations and in the operating room. The relevance of this literature is to provide background information about the performance measures used in the current study, in which operative performance was the primary outcome.

1. Assessment of laparoscopic simulator performance

MISTELS (FLS Simulator)

In the process of developing the MISTELS program, experienced surgeons were asked to review videotapes of operations and list the skills they thought were fundamental to laparoscopic surgery. Another group of experts reviewed the MISTELS modules and completed a global rating scale to determine which of the fundamental skills needed for laparoscopic surgery were represented in MISTELS[11, 18]. During the last decade this simulator and its metrics have been validated extensively. Construct validity was determined by comparing MISTELS performance in more than 200 surgeons with varying laparoscopic experience from 5 countries[10] . Concurrent validity was demonstrated by showing correlation between MISTELS performance and the subjective but widely used In-Training Evaluation Report (ITER) in residents[19], and also with performance in a porcine model of laparoscopic tasks[9]. McCluney et al used a multivariate model to show that FLS simulator scores were highly predictive of operative performance, independent of the level of the surgeon's experience[16]. Inter-rater and test-retest reliability were shown to be excellent (0.99 and 0.89 respectively), and internal consistency (cronbach's alpha) was 0.86. All these meet the standard of reliability needed for high stakes examination[20].

Virtual Reality

Virtual reality is defined as the collection of technologies that allow people to interact efficiently with three dimensional computerized databases by using their natural senses and skills [21]. The minimally invasive surgical trainer – virtual reality (MIST-VR) system was one of the first laparoscopic VR simulators. A collaboration between surgeons and psychologists performed an ergonomic evaluation of the psychomotor skills involved in performing laparoscopic surgery resulted in a toolkit of skills required to perform procedures successfully[21]. Each task is programmed to deliver varying degrees of difficulties to the trainee, and performance is recorded and saved for later analysis of different aspects such as accuracy, time to complete tasks, economy of movement, and others. These output data are referred to as the simulator metrics and have been validated extensively for assessment of basic laparoscopic skills[22]. Gallagher et al found very high internal consistency of the VR performance measures, but a more variable test-retest reliability that ranged from 0.50 to 0.96[23]. One of the main advantages of VR technology is that it can provide objective feedback on the psychomotor performance indicators it measures. One of its main drawbacks is that it is extremely difficult to incorporate realistic haptic (tactile) feedback into these systems. This fact has a crucial effect on the performance of advanced tasks/ skills like laparoscopic knot tying. Botden et al compared basic and advanced tasks on Lap-Sim (a VR simulator) and Pro-MIS, a new laparoscopic simulation system offering a combination of VR technology with physical objects. The difference in terms of sense of realism was very significant in favor of Pro-MIS[24]. This can be attributed to the part of Pro-MIS that acts as a physical simulator – the use of real OR instruments, needle, thread and simulated tissue are far better in terms of sensation that the virtual environment.

2. Assessment of operative performance

Motion Analysis

One method of assessing technical skills is based on dexterity analysis systems. The Imperial College Surgical Assessment Device (ICSAD) is an electro-magnetic tracking system, which consists of a generator and two sensors that are attached to the dorsum of the surgeon's hands. Positional data generated by the sensors is converted to dexterity measures, such as the number, speed and direction of hand movements. Different studies have shown the construct validity of the ICSAD with respect to open and laparoscopic surgical tasks[25-27]. Experienced laparoscopic surgeons are more economical in terms of number of movements, more accurate in terms of target localization, and therefore show shorter path-length.

Another motion analysis system is the Advanced Dundee Endoscopic Psychomotor Trainer (ADEPT), originally designed to select trainees for endoscopic surgery. The motion tracking system here is base on infrared camera, surrounded by infrared light-emitting diodes. Sensors placed on the limb of the surgeon reflect the infrared light and the computer analyses the motion of the surgeon's hand. Validity and reliability of this system have been shown[28, 29]. The two main disadvantages of this system is the inability to use sensors on both arms due to signal overlap, and the lost data whenever the camera and generators become obscure from the sensors.

Data that is generated from these dexterity analysis systems is automatic, objective, and instant, but its main drawback is that it does not provide any information regarding the quality of the procedure performed. In addition, while the assessment can distinguish between surgeons of varying expertise, the data provides little of educational value for surgeons trying to acquire that expertise.

End-Product Analysis

The efficiency of task performance can be misleading in the sense that it does not necessarily include any measure of quality in the assessment. End-product analysis provides important information about the performance quality[30]. Looking at suturing skill as an example, neither the time to complete the task, nor the number of movements or path-length, will teach us anything regarding the quality of the knot. The knot should be inspected and its quality assessed with measured variables like the distance between where the needle was passed and the location of the target. Some errors like an incomplete knot, or one that falls apart when tension is applied, should be considered "critical errors" and should result in failure, no matter how quickly the task was completed. The ideal method of assessment in surgery should incorporate some component of the final product, but outcomes after surgery are often difficult to ascribe solely to surgical technique, and the adverse outcome from poor technique may not be apparent for many years (e.g. cancer recurrence)[31]. For that reason researchers have looked into outcome measures on bench models instead of the clinical arena. Szalay et al assessed the quality of the final product on a 5 point scale after performance of six different bench model tasks[32] and demonstrated construct validity, as well as correlation between final product scores and global assessments of performance. Hanna et al measured the quality of laparoscopically performed knots by using a tensiometer[33], and derived a quality score for knots as an index of knot reliability. Some researchers believe that the combination of these final-product measures with dexterity data would allow to derive proficiency scores, and make the assessment more objective[31].

Error analysis

Ever since the "wake up call" in the report of the Institute of Medicine "to Err is Human" in 1999[34], the literature on the nature of errors and patient safety has grown significantly. In defining the nature of errors, there is an implicit understanding that the error is unintended[35]. Clearly, no surgeon would intentionally commit an error. Errors can be of three types: commission – doing the wrong thing, omission – not doing the right thing, and execution: doing the right thing incorrectly. This understanding helps identifying the source of the error and implies methodology for correction. Seymour et al brought this concept to their new methodology of assessment of OR performance[36]. They defined 8 events associated with the gallbladder dissection from the liver bed as deviations from optimal performance, and created a scoring matrix that allows the evaluator to record whether one or more of these events has occurred for each minute of a video-taped procedure. They showed a high inter-rater reliability of 0.9, and established this tool as a valid measure of performance assessment in the OR. The main drawback to the approach is that it is cumbersome, requiring minute-by-minute analysis.

Checklists and Global Scales

Checklists and global rating scales are the most frequently used methods to assess technical or procedural skill, and a global rating scale (GOALS) was used in the present study to evaluate operating room laparoscopic performance. A checklist consists of a list of procedural steps that are either performed correctly by the examinee or not. Checklists have the advantage of turning the examiners into observers, rather than interpreters of behavior, thereby removing the subjectivity of the evaluation process[37]. Nevertheless, they can be quite cumbersome in the sense that they mandate the full attention of the evaluator to the many little details of the trainee's performance. A second method employs Global Rating Scales (GRS) to assess procedures as a whole, rather than breaking them down into procedure-specific steps. Regehr et al showed that a GRS used to assess resident performance on an eight-station bench model examination (OSATS) had higher reliability than task-specific checklists. This GRS also performed well when used to assess live animal operations, without any differences in the performance of trainees seen in the two model formats[14]. The seven domains assessed by the OSATS GRS are not task-specific, and include respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants, and knowledge of specific procedure. Each is scored from 0 to 4, and the marks across the dimensions are summed. The main drawback to the OSATS GRS is that some important aspects of technical skill, such as the ability to dissect in the correct tissue plane and adequately retract, are not assessed. Importantly, fundamental skills in laparoscopic surgery, like the use of both hands and depth perception, are not evaluated.

The set-up of the OSATS examination requires that the examiner be present in real-time to assess the trainee, which is resource-intensive and increases the potential for bias. Another option is to retrospectively watch videos of the procedures, which improves the objectivity of the assessment, but makes the assessment of some of the domains more difficult – e.g. use of assistants, knowledge of the specific procedure or knowledge of instruments. For that reason Dath et al developed the Operative Component Rating Scale (OCRS), a procedure-specific scale that assesses each procedural component separately on a 5 point Likert scale[38]. They assessed senior surgical residents that were videotaped performing two advanced laparoscopic procedures on pigs, reporting high inter-rater reliability – close to the 0.80 level needed for high stakes evaluations, such as credentialing decisions. The advantages of assessments based on videotaped procedures are the ability of the raters to view the procedure on their own time, and the ability to shorten the assessment time by using fast-forward. But this might be a drawback in the sense that self-editing of the video may cause loss of valuable information for the performance analysis. Scott et al compared the assessment of performance of laparoscopic cholecystectomy between edited videotapes and direct observations[39]. They tried to avoid the pitfalls of editing the videotapes by using the initial few minutes of three predefined parts of the procedure as samples of the residents' skills. However, they found very low inter-rater reliability of the videotapes (0.28), as well as poor correlation with the operative assessments ($r=0.33$). Thus there remain advantages and drawbacks to both real-time and video-taped assessment approaches.

GOALS- Global Operative Assessment of Laparoscopic Skill

The review above highlights some of the limitations in methods to assess laparoscopic performance in the operating room. The OSATS GRS is probably the most widely used instrument, with good evidence for its reliability and validity, as well as ease of use and applicability to both live and video-taped surgery. However, it does not assess unique skill domains pertinent to laparoscopic surgery, and open surgical skills do not imply good laparoscopic skills- in fact, the most experienced open surgeons may have the most difficulty transitioning to laparoscopy, as was demonstrated recently by Vickers et al for prostate cancer[40]. In response to this need, our group developed a new instrument, the Global Operative Assessment of Laparoscopic Skill (GOALS). Like the OSATS GRS GOALS is a generic global rating scale not specific for a particular procedure and covers domains such as efficiency and tissues handling, but adds domains assessing bimanual dexterity and depth perception. Each if the five domains is scored from 1 to 5, with anchor descriptions at 1,3 and 5; the scores are summed to give a total score that range from 5 to 25 [Appendix III]. When GOALS was compared to a 10-point task specific checklist assessing dissection of the gallbladder from the liver bed and to a VAS of overall technical competence, measures of reliability were higher for GOALS than for the other measures, confirming what was reported for the OSATS when it was introduced[37].

Although the evaluator needs to be an interpreter, the objectivity of the tool is reflected in its high inter-rater reliability. Evidence for construct validity has been provided by us and by others for dissection of the gallbladder from the liver bed[15], entire LC, laparoscopic appendectomy[41] , and incisional hernia repair[42] . GOALS can also be used to assess laparoscopic skill based on videotaped performances, but raters need to have had some experience with the tool in the OR first[43]. GOALS appear to be the "right mix" of structure and objectivity while allowing the evaluator to incorporate his or her experience for more comprehensive assessment[15].

In fact, GOALS measures more than simple technical skills. Two domains, efficiency and autonomy, incorporate the cognitive part of the competency needed to perform the surgical task safely and accurately. Theoretically, errors in performance can be done even with perfect technical ability, e.g. unrecognizing the anatomy/ pathology and therefore dissecting in a wrong plane/ cutting a wrong structure. Most bile duct injuries were not the result of poor technical skills but rather resulted from incorrect interpretation of the image on the screen[44]. GOALS is attractive in that it enables constructive feedback, related to specific components of technical skills. Residents are able to direct their practice and improve their performance accordingly e.g. resident who has a problem with bimanual dexterity can go back to the simulator and practice a transferring task until proficiency is achieved, then be assessed again in the OR to verify the effectiveness of this training. The utility of this type of feedback is contrasted with assessments based on speed or dexterity analysis (e.g. instrument smoothness, path length) alone for example. Finally, GOALS is a very practical tool simply requiring observation of the procedure and no more than 5 minutes at the end of the case to complete the score sheet.

II – Skill Transfer from the Simulator to the OR

Of course, the ultimate goal of simulator assessment and training is to improve the competency of trainees in the real OR tasks they will encounter in practice. In order to justify the capital investment and cultural changes required to incorporate simulation into surgical residency curricula, it must be demonstrated that skills learned in a laparoscopic simulator transfer to the operating room environment. A prerequisite for these studies is the use of a valid and reliable assessment tool for the outcome -operative performance - as reviewed above. Previous randomized trials and systematic reviews addressing the transferability of skills between laparoscopic simulators and the operating room will next be summarized (Table 1). Strengths and weaknesses of this previous work will be highlighted, in order to better situate and explain the design and interpretation of the present study,

While early work of Derossis et al demonstrated that training on the MISTELS simulator leads to better performance on the simulator[45], in a subsequent study, they investigated whether simulator practice would correlate with improved performance in an animal model. 12 PGY3 residents were randomized to either five proctored MISTELS practice sessions or no practice. All subjects underwent baseline and final evaluations in the simulator and in the animal model. Of note, the same tasks were performed in both models with slight modifications, and the same metrics for assessment were used as well. The performance in the simulator correlated significantly with the performance in vivo. Both groups improved their in vivo score from baseline to final evaluations, but the magnitude of the improvement was significantly greater for the practice group[9].

In 2000, Scott et al, in a well designed study, randomized 27 PGY 2-3 residents to simulator practice group or control[46]. For simulator assessment and training they used a video-trainer with a curriculum that was based on five established laparoscopic drills: Checkerboard, Bean drop, Running String, Block Move and Suture Foam. This is referred to as the Southwestern Simulator. Scores were recorded as the average time necessary for task completion. OR performance was assessed during an elective laparoscopic cholecystectomy using the OSATS GRS. The practice group trained together as a group for at least 30 minutes daily for 10 days, without any specified supervision. At the end of this period, final evaluations were made in the simulator and the OR. Improvement for OR performance was defined as the difference between baseline to final evaluation for each domain, adjusted for baseline score. They did not analyze changes in the total (summed) GRS score. Improvements in OR performance were significantly greater in the trained group in 4 out of the 8 domains of the GRS – respect for tissue, instrument handling, use of assistants, and overall performance. They concluded that intense training improves video-eye-hand skills and translates into improved operative performance in junior residents. Two limitations to this study include: first, the lack of a predefined goal for simulator training to ensure that subjects all obtained the benefits of training, which may take a different amount of time for each subject, and second, the assessment tool for operative performance was not previously validated for this procedure, or any other laparoscopic procedure.

Hyltander et al looked at skill transfer from a VR simulator (LapSim) to the operating room[47]. They randomized 24 medical students undergoing courses in surgery to train on LapSim for 2 hours per week for 5 weeks, or to control. The simulator tasks included camera and instrument navigations, and coordination. OR performance was measured in a porcine model using tasks they designed to resemble those from the simulator: navigation of 30° camera, instrument navigation and combination of the two. Their performance measures included time to complete the task and a subjective evaluation by 4 senior surgeons of the videotaped procedure. The trained group scored better both on the simulator and in the pig model. Again, neither the tasks nor the metrics that were used for assessment of OR performance were previously validated.

VR simulation was used in two other important trials concerning skill transfer. Seymour et al randomized 16 PGY 1-4 to either training on the MIST-VR or control groups[48]. At baseline, innate ability (visuospatial, perceptual and psychomotor) was assessed using a series of validated tests. For training purposes, 4 experienced laparoscopic surgeons performed the task “manipulation & diathermy” on the simulator and set the training goal for the residents. For the assessment of OR performance, the investigators created a checklist of 8 specific errors in the dissection of the gallbladder from the liver bed during elective laparoscopic cholecystectomy (LC), and measured the time taken to complete this dissection. Procedures were videotaped and scored on a minute-by-minute basis using a scoring matrix that recorded whether an error had occurred during each 60-second period. They demonstrated very high inter-rater reliability for this assessment tool. In all error categories but two, more errors were observed in the control than in the training group, and the duration of the dissection was 29% shorter for the trained group. This study is important because, for the first time, training was defined here to a criterion of expert level, and the assessment tool of OR performance showed reliability. Its limitation is in the fact that baseline evaluations of OR performance were not made. MIST-VR was also used by Grantcharov et al, who randomized 16 junior residents to simulator training (ten repetitions of each of the six tasks) or control[49]. Video-taped OR performance was assessed before and after training using a four-domain GRS based on OSATS but significantly modified. This group evaluated the part of the procedure (LC) starting from the point at which clips were applied to the cystic artery and duct and finishing with dissection of the gallbladder from the liver bed. Residents that received VR training performed the operative task faster and with greater improvements in error and economy of movement scores compared to controls.

Korndorffer et al used the MISTELS (FLS) platform to develop a performance-based laparoscopic suturing curriculum. Their purpose was to test the transferability of this curriculum to the OR. The idea of defining a performance-based curriculum was based on their previous work with the Southwestern simulator[50]. Two experts with known proficiency in laparoscopic suturing performed 10 repetitions on the video-trainer model and their mean score was used to define the performance goal. A live porcine Nissen fundoplication model, previously described for other purposes[51], was used to assess operative performance. Scoring was done by a single rater via direct observation, and used the same metrics as for the video-trainer (i.e. single score that rewards speed and penalized errors). 17 residents PGY 1-5 were randomized to training or control groups. Training was fully supervised, and was considered complete when the defined proficiency level was achieved or after 8 1-hour weekly training sessions; all subjects in the trained group achieved proficiency levels after a mean of 2.5 hours. In the final evaluation the trained group performed significantly better on the pig model, both in completion time and in overall score, but their scores were still significantly below the expert scores on the animal model. They suggested that proficiency-based curricula should be further developed.

Schijven et al conducted a case-control study in which they assessed the value of a 4-day LC course by measuring operative performance in LC's in humans[52]. 12 residents who participated in the course were matched to 12 counterparts in terms of their clinical exposure (no more than 4 procedures done previously). The course included both basic and procedural LC in the MIST-VR simulator, as well as progressive participation in actual LC's, first as assistants, and in the last day of the course as primary surgeons under close supervision. OR performance assessment was conducted only on the clipping and cutting of the cystic artery and duct using video – tapes of the procedures. The outcome measures consisted of a combination of a GRS of 2 domains (fluency and carefulness) on a 5-point Likert scale, with metrics that were modified from the VR simulator, and a VAS scoring "judgment" from 1-10. The trained group performed significantly better in all domains. There are a few limitations to the methods of this study: first, group assignment was not done randomly, and second, baseline skills were not assessed in the control group.

Ahlberg et al Sweden randomized 13 inexperienced residents at 9 centers to proficiency-based VR training or control[53]. The impact on performance of the first 10 entire LCs was investigated. For simulator training, proficiency was defined as the median values obtained by 5 expert laparoscopic surgeons in 6 different tasks on the LapSim VR simulator. Assessment in the OR was conducted in the first, fifth and tenth LC for each participant. Procedures were divided into 3 different phases: exposure of the cystic duct and artery, clip placement followed by cutting, and gallbladder dissection. Procedures were video-taped and assessed by 2 blinded observers, using the minute-by minute error scoring matrix previously used by Seymour et al[36],[48]. The VR trained group consistently made fewer errors then the control.

Aggarwal et al assessed the effect of a proficiency-based VR training curriculum on the learning curve of LC in a cadaveric pig model[54]. They randomized twenty novice surgeons to a VR-trained group who completed a training curriculum followed by 3 cadaveric porcine LC's each, and a control group who performed 5 LC's on the same model. For training purposes they used the LapSim VR simulator with tasks and curriculum that they had previously validated, which includes a simulated LC operation as well as other drills[55]. Each cadaveric LC was assessed using their previously validated video-motion analysis system (ROMIVAS)[56], and scored with the

OSATS-GRS. 10 experienced surgeons completed two cases each in order to define the proficiency levels on the cadaveric porcine model. The trained group outperformed the control on all measured parameters. In their fifth LC, the control group reached statistical equivalence to the performance of the third LC in the trained group, in terms of the dexterity measures. The video rating scores, though, remained significantly higher in the trained group, even in these last cases. One of the strengths of this study is the use of a combined approach to the performance assessment – dexterity measures, and the OSATS-based GRS. It is interesting to see, though, that in terms of the dexterity measures, the trained group reached statistical equivalence to the experts' levels by their third case, and even the control group reached this equivalence for path length by their fifth case (ceiling effect). Only the GRS measures remained significantly different in favor of the trained group at their last cases.

Two systematic reviews of skill transfer are available. The Australian surgical technology assessment group ASERNIP-S published a review of studies assessing skill transfer after surgical simulation training (physical or VR)[57]. They only included studies that assessed operative performance on human subjects either in the OR assessing LC, or in the endoscopy suite during colonoscopy/ sigmoidoscopy. The 5 studies assessing LC included 3 small RCT's[46, 48, 49, 52] (59 subjects randomized in total), and one case-control study[52]. The fifth study[58] was consisted of preliminary data from Scott et al[46] presented at the forum of the American College of Surgeons. The review emphasized the fact that there were large variations in the assessment methods used for operating room assessment: Scott et al assessed the performance of the entire procedure[46], Schijven et al assessed the clip and cut portion[52], Grantcharov et al assessed the part from the clip and cut and included the dissection from the liver bed[49], and Seymour et al assessed gallbladder dissection only[48]. There was also substantial variability in the type of simulators, which included both virtual reality (MIST-VR) and a physical trainer (Southwestern), the training tasks used (all tasks or selected tasks for each simulator) and the training methods (supervised or not, to a pre-defined proficiency level or by time). Despite concluding that "skills acquired by simulation-based training seem to be transferable to the operative setting", the authors cautioned that "the studies were of variable quality and did not use comparable simulation based training methodologies, which limited the strength of the conclusions". The authors called for more studies to be performed to strengthen the evidence-based and establish the role for simulation in training curricula.

A systematic review investigating the effectiveness of VR simulator training (but not video-trainers) for laparoscopic surgery is also available[59]. This includes 23 randomized studies in which the effectiveness of VR training was compared with physical-simulator training, to no simulator training or to standard laparoscopic training. A variety of OR outcome measures both in animals and humans were included. Only three trials had adequate methodological quality in terms of generation and concealment of allocation, blinding and follow-up [54, 60]. Only one trial mentioned patient-related outcomes, conversion rate to open LC[53].

For both complete novices and trainees with limited laparoscopic experience, VR training was found to improve upon standard laparoscopic training.

III- Simulator training in basic laparoscopic skills

Type of simulator: VR or physical (video-trainers)

Several studies have compared VR simulators to physical simulators (Video-Trainer or VT) in terms of their influence on the acquisition of laparoscopic skills. Hamilton et al assessed 50 residents on both systems and than randomized them to practice on either the VR or VT system[61]. Each group showed the greatest improvement on the simulator they trained on, but skills were also transferable from one simulator to the other (more so for the VR group). Operative performance however improved significantly only in the VR group. Munz et al took 24 complete novices and randomized them to 3 groups – VR, VT and control. They assessed their performance on the box-trainer before and after training, using motion analysis and error scores[62]. The VT group performed better than the control in all parameters, whereas the VR group did better on some of the parameters. Madan and Frantzides randomized 65 novices to training either on VR, VT, both or control[60]. Assessment was performed on a pig model with 4 tasks designed to represent portions of commonly performed laparoscopic procedures. Only the group that trained on both simulators showed significant difference from the control. To date, there is little evidence to consistently support the superiority of one system over the other[57, 59, 62]. This is consistent with non-laparoscopic simulation studies in endourology[63] and microsurgery[64], where the superiority of a "high fidelity" environment has not been demonstrated compared to "low-fidelity" bench-models, as long as the pertinent steps of the procedure using the same skills can be practiced. High fidelity models are much more expensive, and in laparoscopic surgery they are always in VR simulators. The lack of haptic (tactile) feedback, or its unrealistic nature in most of these simulators, keeps their level of fidelity less than high. In essence, for basic laparoscopic skills acquisition, low fidelity models in the form of video-trainers like FLS seem to be as good as their VR counterparts, and certainly cost less. The present study evaluated, for the first time, the transfer of training for the manual skills component of the FLS program, the most widely used VT trainer.

Simulator training strategies

For the past century or so, North American surgical training has relied on the apprenticeship model. Training ended after a predefined time (generally 5 years), during which a trainee spent long hours in the hospital (e.g., on call every 2nd night, hence the term "resident") and was given an increasing degree of responsibility and independence. With the long hours, it was presumed that the time allotted would ensure a sufficient volume of cases to enable competency. This approach has been challenged for a variety of reasons, including the changes in surgical technology, focus on medical errors, cost and scarcity of OR time, and resident work-hour restrictions. However, there is also a sound psychological basis for a paradigm–shift in surgical skills training from time-based or number of cases-based training to one that is proficiency-based[65]. That is, since individual surgeons differ in their innate visuospatial, perceptual and psychomotor abilities, it follows that they would require different amounts of practice time to gain automaticity for their technical skills. In other words, as for motor skills in general, the rate of learning to expertise level differs between surgeons and benefits from deliberate practice.

The core of this new paradigm is the establishment of proficiency criteria. Benchmark levels should be established from objectively assessed performance of experts for each task. The experts should reflect a representative sample of the proficient population. Seymour et al were the first to use this approach in their trial of skill transfer from MIST-VR to cholecystectomy[48]. In 2007 Ritter and Scott designed a proficiency-based skills training curriculum for the FLS program[66]. The training curriculum in our study was modified from their design. Proficiency criteria are incorporated more and more as training goals for basic laparoscopic skills on different platforms[54, 67, 68].

Regarding practice strategies, task sequence should progress according to degree of difficulty (easier to harder), with smooth transitions between tasks and feedback proximate to performance. Distribution of practice has also been studied and interval practice consistently showed more improvement than massed practice. The likely explanation is that skills being learned have more time to be cognitively consolidated between practices[65].

The importance of constructive feedback to the trainee cannot be overemphasized. Grantcharov et al studied 16 residents performing two laparoscopic cholecystectomies in the OR within a two-week period[69]. They were divided into 2 groups: the intervention group received a 60-minute structured constructive feedback session based on video-taped assessments after their first procedure. The other group served as control. Performance assessment in all procedures was done with a GRS. Residents who received feedback demonstrated significantly greater improvement in their time, error and economy of movement scores from the first to the second case.

In summary, we have seen in this review that the majority of studies that evaluated skills transfer or generalization from simulator to real OR have used a VR system as their simulator platform. The only study that used a physical simulator (video-trainer)[46] used time-based group training, and an assessment tool for OR performance that has never been validated for laparoscopic surgery. These limitations brought our group to design this randomized controlled trial.

In this trial we evaluated, for the first time, skills generalization in junior surgical residents after a proficiency-based training curriculum, in the most widely used laparoscopic simulator – FLS. For OR performance assessment we used, again for the first time, a GRS that has been validated specifically for laparoscopic surgery – GOALS.

SECTION C -

FLS SIMULATOR TRAINING TO PROFICIENCY IMPROVES LAPAROSCOPIC PERFORMANCE IN THE OPERATING ROOM –

A RANDOMIZED CONTROLLED TRIAL

I- Contribution of Authors

Gideon Sroka M.D.: took part in the study design, recruited and trained the study subjects, managed and analyzed the data and wrote the manuscript.

Liane S. Feldman M.D.: designed the study, performed intraoperative assessments, supervised data management and analysis, edited the manuscript.

Melina C. Vassiliou M.D.: designed the study, obtained IRB committee approval.

Pepa A. Kaneva M.Sc.: participated in data management and analysis.

Raad Fayez M.D.: participated in recruitment and randomization of subjects.

Gerald M. Fried M.D.: designed the study, performed intraoperative assessments, helped with data analysis and interpretation, edited the manuscript.

II- Introduction

In the past two decades since their introduction, video-laparoscopic techniques have been widely adopted by surgeons from various disciplines including general and thoracic surgery, urology and gynaecology. Residency programs are facing the educational challenge of teaching unique skills required to safely perform laparoscopic surgery. In addition, the renewed emphasis on patient safety, ethical issues inherent in “practicing on patients” and pressures to improve operating room efficiency have all become increasingly relevant. The traditional “see one, do one, teach one” approach is clearly not appropriate in this new environment.

Surgical training has a long tradition of incorporating simulation in the guise of cadaver, animal and bench top models to develop and practice open procedural skills. The opportunity to acquire the novel skills required for laparoscopic surgery in an environment that is efficient, effective, and does not jeopardize patient safety is also very appealing. The MISTELS program (McGill Inanimate System for Training and Evaluation of Laparoscopic Skills), based on a physical laparoscopic trainer box, was developed with the aim of standardizing the teaching and evaluation of fundamental laparoscopic skills in a safe environment [70]. MISTELS is an inexpensive, portable and flexible system, and has been extensively validated

[10, 20]. It has been shown that performance in the simulator improves progressively with practice [45] and correlates with operating room performance [16]. MISTELS was incorporated as the manual skills component of the Fundamentals of Laparoscopic Surgery (FLS) program developed by the Society of American Gastrointestinal Endoscopic Surgeons (SAGES) and endorsed by the American College of Surgeons (ACS).

While the FLS simulator has become a *de facto* standard for evaluation of technical skill in laparoscopy [71], the question remains whether or not skills gained from training on the FLS simulator translate into better performance in the operating room (OR). Although transfer of learning from this simulator and other simulators to the operating room has been evaluated [9, 46-49, 52, 53, 72], the studies have been hampered by lack of a standardized simulator and curriculum, inconsistent proctoring during training, and especially by the absence of a validated measure of laparoscopic operating room performance [57]. The Global Operative Assessment of Laparoscopic Skills (GOALS) was developed to evaluate intraoperative laparoscopic skills by direct observation. It was initially validated during dissection of the gallbladder from the liver bed, and has been shown to be a reliable and valid measure of technical skill in the clinical setting [15]. The aim of this study was to assess the transfer of skills acquired by novices trained to proficiency on the FLS simulator to operating room performance as measured by GOALS.

III - Methods

Nineteen general surgery residents in post-graduate years (PGY) 1 to 3 at McGill University participated in this Research Ethics Board approved study (Project **A03-E06-04A***)*. After appropriate informed consent, all participants underwent baseline FLS testing, and were assessed in the OR using GOALS during dissection of the gallbladder from the liver bed during elective laparoscopic cholecystectomy. Those with GOALS scores ≤15 (n=17) were randomly assigned to training (n=9) or non-training (n=8) groups. The training group used the FLS simulator in a supervised proficiency-based curriculum, based on recommendations by Ritter and Scott [66] . Both groups continued their regular residency training and subjects were asked to document their clinical laparoscopic experience throughout the study period. At the end of the study period (mean time between evaluations was 145 days) subjects were assessed again on the simulator and in the OR, using the same metrics. Evaluators of simulated and clinical performance were blinded to training group and to one another. *Figure 1* summarizes the flow of participants through the study.

Ethical Issues

Participants were assured that their evaluation would not change the way the procedure was done or the quality of the care patients received. Evaluations were restricted to dissection of the gallbladder from the liver bed and the attending surgeon could take over at any time. The data gathered were coded and all reporting was confidential and did not impact the resident's official evaluation. Participants could choose to withdraw at any point during the study and they were made explicitly aware of this at the time of informed consent. In order to mitigate the potential ethical dilemma associated with restricting training opportunities thought to be beneficial from one group of residents, FLS training was offered to all those randomized to the non-training group at the end of the study period.

Randomization

Randomization was done by drawing an assignment from a box with a 50% chance of being in the training or non-training group, and was performed by an investigator not involved in the process of training and evaluations. Participants were asked to keep their randomization status confidential for the study period.

Simulator performance metrics

Laparoscopic proficiency training was carried out using the manual skills portion of the FLS program [11], which includes a CD-ROM of didactic material and the 5 MISTELS tasks that have been previously described in detail [20, 45, 68]. Briefly, the simulator consists of a trainer box with an opaque cover, a built-in camera and 2 trocars. It can be attached to any monitor with an s-video connection. The 5 tasks include Peg transfer, Circle cut, Placement of a ligating loop, and simple suture tied with extra and intra-corporeal techniques [73]. All participants went through an orientation, and viewed the video tutorial for all five tasks. Baseline scores were calculated after the first iteration of each task by an experienced FLS proctor using the standard FLS metrics. An overall summative evaluation (pass or fail) was provided after the completion of all tasks. The metrics reward efficiency (speed) but also penalize errors; higher scores indicate better performance. Final evaluations were performed in the exact same way as at baseline.

Intraoperative laparoscopic performance assessment

Intraoperative laparoscopic performance was evaluated with the GOALS. This global rating scale measures performance in five domains, three of which are specific for laparoscopic surgery (depth perception, bimanual dexterity and tissue handling), and two that are more generic (efficiency and autonomy). Each domain is scored on a Likert scale from 1-5, with anchoring descriptions at 1, 3, and 5. The total score ranges from 5-25, with higher scores indicating better performance. This tool has been shown to be reliable, valid, and feasible for evaluation of dissection of the gallbladder from the liver bed. We previously reported that junior residents (PGY 1-2) achieved a score of 12 [95% CI 11-13], compared to 17 [95% CI

14-21], for intermediate level residents (PGY 3-4) and 22 [95% CI 20-24], for more experienced surgeons (PGY 5+) [16]. All participants were evaluated in the OR at baseline and at the end of the study by the attending surgeon and trained evaluators blinded to their randomization status. The attending surgeon also evaluated the perceived difficulty of the dissection using a visual analog scale (VAS) from 0 (easiest) to 10 (most difficult).

Setting

The study took place within an accredited general surgery residency program. FLS training took place in the Steinberg-Bernstein Centre skills lab at the Montreal General Hospital. In addition to proctored sessions, the training group was given a key card enabling unscheduled access to the skills lab at any time. Operating room assessments took place at 2 McGill University teaching hospitals.

Training curriculum

The training goal in the simulator was based on a study by Ritter and Scott [66]. They used their own expert performance to set a specific time goal with “allowable” errors for each task. Proficiency was defined as performing each task within the specified time and error goals on at least two consecutive repetitions (with 10 additional non-consecutive repetitions for tasks 1 and 5).

Scott et al have also shown that this proficiency based curriculum is feasible for training novices, and allows sufficient skill acquisition for FLS certification in a mean of 9.7 hours of training [74]. The criteria were revised slightly in our study in that proficiency time for task 1 (peg transfer) was increased from 48 to 60 seconds, a more practical goal for junior residents, but still above the score required to pass the task. Over the study period participants from the training group practiced both under supervision and independently. Proficiency was confirmed by an experienced proctor. Subjects in the training group were encouraged to practice on the FLS simulator in their free time, and decided when to be supervised by the proctor. They were asked to document the number and time of training sessions. For each participant, the study period was considered complete when proficiency testing was requested by the subject and confirmed by the proctor for the training group, and after at least 6 weeks for the non-training group.

Statistical analysis

The primary outcome was the change in operating room performance measured by GOALS. We determined the required power of the study using previously collected data on GOALS scores in 76 surgical residents. This demonstrated a mean score of 12.3±2.8 for PGY1-2 and 17.5±3.6 for PGY 3-

4. We defined 5 as a clinically meaningful difference in score. We calculated that 7 subjects in each group would give a power of 80% to detect a difference of 5 points in GOALS scores with an alpha of 0.05. Statistical analysis was conducted using SPSS version 11.0 (SPSS Inc. Chicago, IL). Student's t test was used to compare mean FLS and GOALS scores in the training and non-training groups. Paired t tests were used to compare baseline and the final performance within each group. Multivariate analysis was used to assess the effect of simulator training on GOALS score after adjusting for baseline GOALS score. Data are expressed as Mean±SD. Demographic data (Table 2) are expressed as median (IQR). $p<0.05$ was considered statistically significant.

IV- Results

Of the 17 subjects who were randomized, 16 completed their final evaluations, 8 in each group. The groups were similar at randomization (Table 2). Figure 2 shows the change in FLS simulator scores over the study period. No differences were found for the baseline FLS scores in the trained and non-trained groups. At the final evaluation, FLS scores increased in the trained group to 95.1±4, compared to 60.5±23 in the non-trained group ($p=0.004$). At baseline, no participant had a score above the level required for FLS certification; at the second evaluation, all trained subjects would pass

FLS, whereas only 3 of the 8 non-trained subjects had a passing score. All participants in the trained group met the predefined proficiency criteria after mean training time of 450 min (in 9 separate sessions), out of which 150 min were proctored, and the rest unsupervised.

Figure 3 summarizes the change in operating room performance as measured by GOALS score, for the two groups. Baseline operating room performance was similar ($p=0.47$). Participants in the non-trained group improved their performance by 1.8±2.1 points from 12.0±1.8 to 13.8±2.2 ($p=0.04$), a clinically insignificant magnitude. In contrast, the trained group improved significantly by a mean of 6.1±1.3 points from 11.3±2.0 to 17.4±1.9 ($p=0.0005$ vs. control; $p<0.0001$ vs. baseline). After adjusting for baseline GOALS score and gender using multivariate analysis, only group allocation remained significantly associated with higher final GOALS score. Of the 5 individual domains evaluated by GOALS, simulator training was associated with greater improvements in the laparoscopy-specific domains (bimanual dexterity, tissue handling, depth perception) compared to the more generic domains (efficiency and autonomy) (Table 3). There was no difference in the attending surgeon's assessment of the difficulty of the dissection for the trained and control groups at the baseline (2.5 vs. 3, $p=0.65$) or final evaluations (4.5 vs. 2.5, $p=0.15$).

V- Discussion

Technical skills are essential to the practice of surgery. There is great interest in the potential for teaching fundamental skills to novice surgeons in the simulator laboratory instead of the OR, both to address patient safety concerns and improve OR efficiency. Repetitive, goal-directed practice of psychomotor skills prior to the operating room may allow some of these skills to become automated. Automaticity implies that when the trainee is then in the clinical environment, his or her attention can shift from the required manual skill to focus on the cognitive aspects of the tasks he or she is facing in the OR, namely, perception (recognition of the anatomy/pathology) and forecasting (looking ahead to the next step) [75]. However, incorporation of simulation into surgical training curricula also requires significant effort, time and money, and will only be fully established when its educational effectiveness is proven. This single-center randomized single-blinded trial demonstrates that training to proficiency using the FLS simulator curriculum improved operating room performance during laparoscopic dissection of the gallbladder from the liver bed in novice surgeons compared to standard residency training. Training to proficiency required an average 2.5 hours of supervised training and 5 hours of unsupervised practice on the simulator. Based on previous work, the magnitude of the difference in the improvement in operating room performance between the two groups is clinically relevant, and is the equivalent of approximately two years of clinical training [16].

Our results are consistent with other randomized studies demonstrating the transferability of skills learned in a variety of laparoscopic simulators to performance of laparoscopic cholecystectomy or its component steps in humans. In the only previous study using a physical box trainer, Scott et al. [46] randomized 27 residents to a training group, who had ten 30 minute training sessions on the Southwestern Guided Endoscopic Module, or to a control group. Performance of laparoscopic cholecystectomy was measured using global rating scale domains modified from the OSATS, originally designed and validated for open bench model performance [14]. The trained group showed greater improvements from baseline in 4 of the 8 domains. Two other studies confirmed transferability from a virtual reality psychomotor simulator (MIST-VR) to the operating room. Seymour et al. [48] randomized 16 residents to MIST-VR training until a preset criterion level was reached, then assessed dissection of the gallbladder from the liver bed using a novel error scoring system which had not been previously validated and not performed at baseline. Fewer errors were made by the trained group. Grantcharov et al. [49] randomized 16 residents to 10 repetitions of the MIST-VR tasks. Subjects were assessed during clipping of the cystic duct and artery and dissection of the gallbladder from the liver bed during laparoscopic cholecystectomy using an error score and economy of motion score, also modified from the OSATS. The simulator-trained group showed greater improvement from baseline and was slightly faster compared to controls. A systematic review of skills transfer after surgical simulation training concluded that for laparoscopic cholecystectomy, subjects who received simulator training prior to operating room assessments performed better than subjects without such training. The authors emphasized the need for more standardized patient-based performance measures: there were large variations in the part of laparoscopic cholecystectomy that was evaluated, as well as in the different metrics that were used for the assessment. Several simulators were used, with diverse platforms, tasks, performance metrics and practice goals (time-based or criterion-based) [57]. None of the previous studies used a patient-based assessment that had been previously validated *specifically* for laparoscopic surgery.

The present study addressed several of these concerns. The FLS simulator was designed to build the fundamental skills required for laparoscopic surgery in novice surgeons. It is an inexpensive skills based simulator that does not model any one specific procedure, and as such has been used in general, urologic and gynaecologic surgery curricula [76]. The FLS curriculum is standardized and readily available at relatively low cost [73]. The performance metrics have been extensively validated with a passing score defined to differentiate competent from non-competent laparoscopic surgeons [68]. FLS certification, which requires passing a cognitive component as well as the manual skills test, will be required for general surgery board certification beginning in 2010 [12]. In the present study, baseline average score (44) was at the level expected for junior residents without simulator training [16]. The criteria used to define proficiency were also evidence-based [66]. The advantage of a proficiency- rather than a time- or number-based goal relates to the observation that people learn at different rates. Using a proficiency goal ensures that this variability decreases with training, as evidenced by the narrowing of the standard deviation in scores from the baseline to final FLS assessment in the trained group. The time to achieve this proficiency level ranged from 5 to 12 hours. Practice was distributed, with each session lasting a mean of 45 min and no more than one hour, with no more than one practice session per day. The session began with a proctored part and definition of what needs to be worked on, followed by unsupervised practice.

We used a patient-based global rating assessment tool (GOALS) that is reliable and valid when used to evaluate laparoscopic dissection of the gallbladder from the liver bed [15], and is strongly correlated with FLS score [16]. Others have provided evidence for construct validity for the GOALS when used to evaluate an entire laparoscopic cholecystectomy as well as laparoscopic appendectomy [41]. We chose dissection of the gallbladder alone to enable participation of junior-level residents, for whom simulation training is expected to provide the greatest benefit. The availability of GOALS data for specific levels of residency training allowed for a sample size calculation to be made based on a clinically relevant difference, since an improvement by at least 5 points would represent the difference between novice and intermediate level residents in regular clinical training. Requiring a GOALS score below 16 avoided the ceiling effect seen for this relatively straightforward operating room procedure. Looking separately at the 5 domains included in the GOALS score, one can appreciate that FLS simulator training targeted the laparoscopic-specific domains of depth perception, bimanual dexterity and tissue handling. Our intervention had little to do with the cognitive part of the procedure, and it was not unexpected that the autonomy and efficiency domains were not affected by simulator training. An additional strength of the study was that the operating room assessments were blinded to the training status of the participants.

The study has several limitations. There was a longer time between the assessments in the simulator-trained group compared to controls, time during which the participants continued their usual training. The time difference was related to the logistical issues inherent in asking the subjects to practice on the simulator in addition to their usual duties, with the simulator lab often located at a different hospital. Our concern about the possibility of bias led us to measure the correlation between this time difference and the main outcome measure, improvement in GOALS score, which was low (Pearson correlation coefficient = 0.13) and nonsignificant ($p=0.63$) (Figure 4). For the training group, the improvement in GOALS score is high, no matter how long the time between the evaluations, as opposed to the control group, where the GOALS difference was variable but also not correlated with the time difference. Despite the discrepancies in time, the subjects reported similar (low) participation in laparoscopic cases during the study period. This may be explained by the fact that these junior residents were not necessarily rotating through general surgery services during the study period. The one case difference between the groups (4.5 vs. 3.5 in favour of the training group) does not seem significant clinically. In addition, the improvement in operating room performance seen in the trained group is greater than expected from usual training alone, as it was similar in magnitude to that reported between a junior and intermediate level resident. The groups were also different in their gender composition, although the difference was not statistically significant. Multivariate linear regression analysis demonstrated that only group allocation, and not gender or baseline GOALS, was significantly associated with GOALS improvement. An additional limitation relates to the small sample size, although this was supported by the sample size calculations and is consistent with previous studies in this area. Finally, this is a single centre study, in the centre where the simulator and all the performance metrics were developed.

In conclusion, we found a statistically and clinically significant improvement in operating room laparoscopic performance in junior residents after proficiency based simulation training compared to un-trained controls. This was achieved using an inexpensive widely available video-trainer after 7.5 hours of simulator training.

SECTION D - CONCLUSIONS

"The overwhelming need facing all surgeons is to improve patient safety and simulation may play a major role in advancing that goal"

The American College of Surgeons

Educational retreat, 2001 [77]

Laparoscopic surgery is very different than open surgery, in terms of the skill set needed for its safe performance. The hand-eye coordination, the need to perform tasks via a two dimensional screen, the increased manual dexterity to compensate for the use of long instruments, the fulcrum effect of the abdominal wall that brings the tip of the working instrument to the opposite direction of its holding hand, and the lack of tactile sensation - all contribute to the difficulty in the acquisition of basic laparoscopic skills. Due to restrictions on working hours, surgical residents spend less time in the hospital and in the OR. At the same time the laparoscopic technique is being used more, and in most of the abdominal operations benefits have been established for the technique: shorter length of hospitalization, better physiologic response to surgery[78-80], and improved postoperative gastrointestinal function[81]. Adding the enormous cost of OR time dictates that basic laparoscopic skills be taught outside the OR. Simulation based training is attractive since it provides a standardized, predictable educational experience, in which errors can be committed without jeopardizing patient safety and where constructive feedback and deliberate practice can occur[3].

The results of our study are consistent with this framework. Deliberate, proficiency-based practice in a widely available, low-fidelity video-trainer was associated with a clinically significant improvement in operating room laparoscopic performance in junior residents, compared to un-trained controls. Although our results are not entirely unexpected, they are certainly not intuitive, since the nature of the tasks performed to proficiency in the simulator are different than the operative task of gallbladder dissection from the liver bed. The FLS training program is not procedure-specific and would not be expected to improve the cognitive part of the performance of LC. Nevertheless, the level of complexity of the OR task that was assessed is relatively low, thus allowing for the technical skills component to be the most influential on the total score. This was seen when the GOALS domains were analyzed separately, demonstrating that the effect of training had the greatest effect on the laparoscopy-specific domains – depth perception, bimanual dexterity and tissue handling. This effect was not apparent on the general domains of efficiency and autonomy.

Researchers have looked at skill transfer when assessing similar tasks in the simulator and the OR. This is easier done on a VR simulator with modules that simulate specific operative steps, and whole procedures, than on a low fidelity physical simulator like the FLS. In that sense, our methodology is similar to the study by Scott et al[46] who used another physical simulator – the Southwestern Guided Endoscopic Module. Lucas et al have shown that surgical skills acquired on a VR laparoscopic simulator with LC training are not procedure specific but improve overall surgical skills that can be transferred to another procedure – laparoscopic nephrectomy[82]. Gallagher et al have made the distinction between Skills Generalization where the trainee learns fundamental skills that are crucial to completion of the actual operative task (as in the FLS simulator) and Skills Transfer that refers to a training modality that directly emulates the task to be performed in vivo (as in higher fidelity models – VR)[65] .

The magnitude of the improvement we demonstrated for the trained group, as measured by the GOALS, was surprisingly high. The GOALS score achieved by the simulator-trained group of junior residents (17) was similar to that of the average PGY 3-4 resident in our previously collected educational database. This implies that a simulator curriculum that is relatively easy and inexpensive (7.5 hours of simulator training – 1/3 of which is supervised) allows for attenuation of the steepest part of the learning curve. This may have training benefits beyond the simple technical skill, allowing for simulator-trained residents to enter the OR able to focus on the cognitive parts of the procedure rather than simple dexterity (e.g. "where is the best anatomic plane of dissection?" rather than "how do I get this instrument to the target?"). This construct is consistent with the evidence that motor skills expertise is defined by the presence of automaticity manifested by the ability to "dual-task" [83].

In conclusion, the results of this trial are the first to demonstrate that skills acquired by junior residents in the FLS simulator transfer to the operating room. The clinically significant magnitude in OR performance was obtained after a relatively short period of FLS training, about one third of which was proctored. FLS simulator training should be incorporated into the curricula for junior residents.

SECTION E – TABLES AND FIGURES

Table 1: Summary of randomized trials investigating skill transfer from the simulator to the OR

LC: laparoscopic cholecystectomy
VR: virtual reality
GRS: global rating scale

Trial	Subjects	Simulator Type	Simulator Tasks	Training Method	OR Task	Model	OR assessment
Fried [9] **1999**	12 PGY 3	MISTELS (Physical)	7 Original Tasks	5 weekly Sessions	Same as Simulator tasks	Porcine	Same as simulator (Time and error score)
Scott [46] **2000**	27 PGY 2-3	South-Western (Physical)	5 Drills	10 daily sessions of 30 min	LC (entire procedure)	Human	GRS (modified OSATS)
Hyltander [47] **2002**	24 Med Students	LapSim (VR)	Camera & instrument Navigation	5 weekly Sessions of 2 h	Camera and instrument Navigation	Porcine	Time and a rating scale
Seymour [48] **2002**	16 PGY 1-4	MIST-VR (VR)	Manipulation & Diathermy	Expert Level	LC (part of procedure)	Human	Errors; time
Grantcharov [49] **2004**	16 Junior Residents	MIST-VR (VR)	All 6 Tasks	10 repetitions of all tasks	LC (part of procedure)	Human	GRS (modified OSATS); time
Korndorffer [72] **2005**	17 PGY 1-5	MISTELS (Physical)	Intra-Corporeal Stitch	Expert Level	Nissen Fundoplication Model	Porcine	Same as simulator (Time and error score)

Ahlberg[53] **2007**	13 PGY 1-2	LapSim (VR)	6 tasks	Expert level	LC (entire procedure)	Human	Errors; time
Aggarwal [54] **2007**	20 PGY 1	LapSim (VR)	All 7 Tasks	Expert Level	LC Model	Porcine	Motion analysis; GRS (modified OSATS)

Table 2. Comparison of simulator trained and control groups at enrolment and during the study period. Data expressed as median (IQR).

	No simulator training (n=8)	Simulator Training (n=8)	p
PGY 1/2/3	6 / 2 / 0	5 / 2 / 1	0.58
Age (years)	27 (27-28)	27 (26.5-28.5)	0.85
Gender (Male/Female)	3 / 5	6 / 2	0.13
Hand Dominance (Right / Left)	7 / 1	7 / 1	1
Time between baseline and final evaluations (days)	113(40-167)	162(100-256)	0.13
LC performed as Primary during study (number)	3.5 (2-5)	4.5 (3-7)	0.21
LC participated as Assistant during study (number)	4.5 (4-6)	4.5 (3.5-8)	0.92
Other laparoscopic cases performed/participated during study (number)	2.5 (2-3.5)	2.5 (1-3.5)	0.75

LC, laparoscopic cholecystectomy

Table 3. Comparison of the difference (mean ± SD) in operating room performance in the domains assessed by GOALS from baseline to final assessment after simulator training compared to controls. Each domain is scored from 1 (worst) to 5 (best) and the results summed to get a total score.

	No simulator training (n=8)	Simulator Training (n=8)	p
Depth perception	0.5 ± 0.8	1.25 ± 0.7	0.08
Bimanual dexterity	0.5 ± 1.1	1.25 ± 0.6	0.04
Efficiency	0.4 ± 1.1	1.13 ± 1.0	0.24
Tissue handling	0.3 ± 0.7	1.13 ± 1.0	0.04
Autonomy	0.3 ± 1.0	0.6 ± 1.1	0.58
Total score	1.8 ± 2.1	6.1 ± 1.3	0.0003

Figure 1: flow of participants through the study.

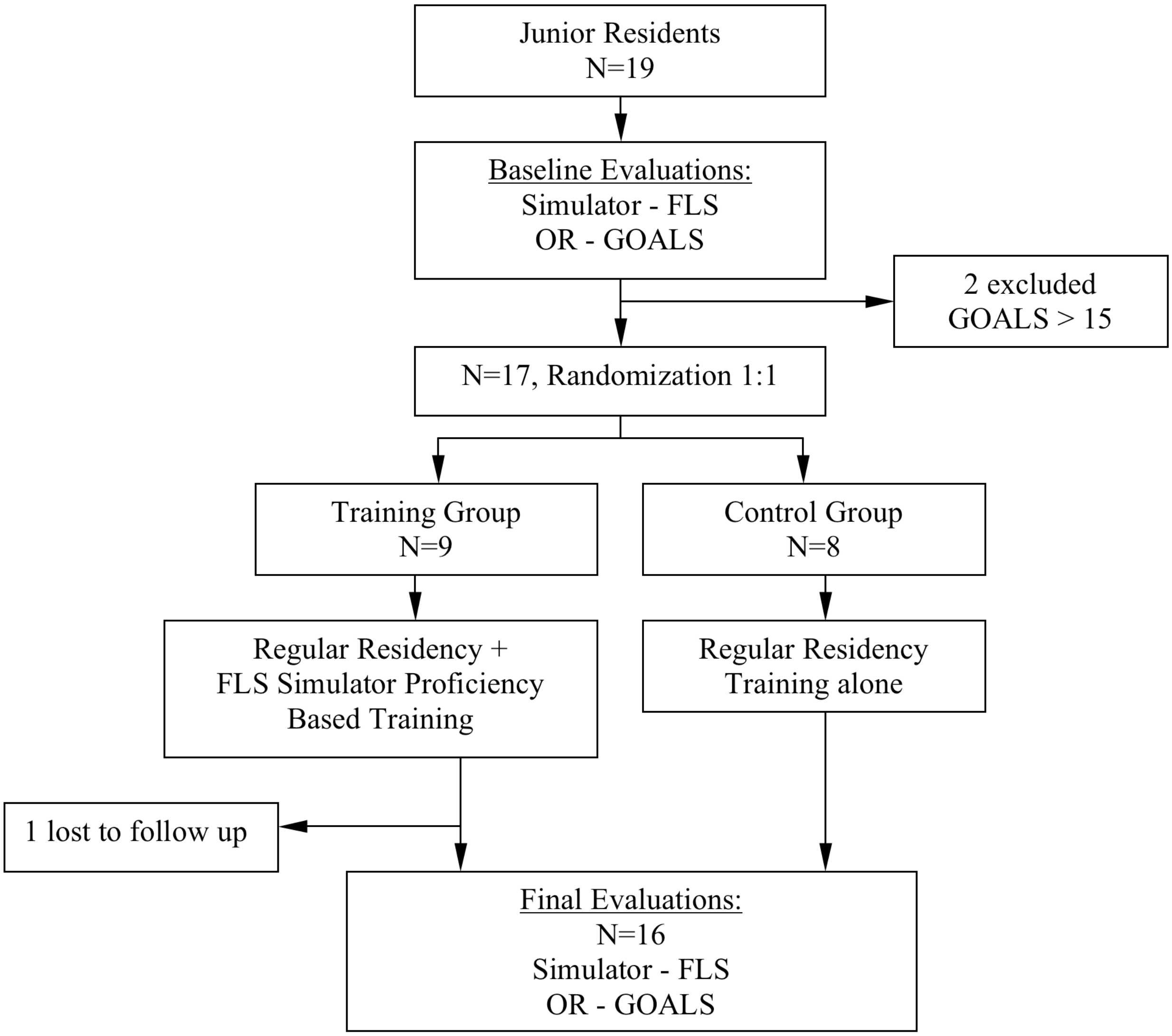


Figure 2: FLS Simulator performance at baseline and at the end of the study period. The groups were similar at baseline (p=0.27). Both groups were improved at the final assessment but the trained group improved more and had a higher score than the control group (p=0.004).

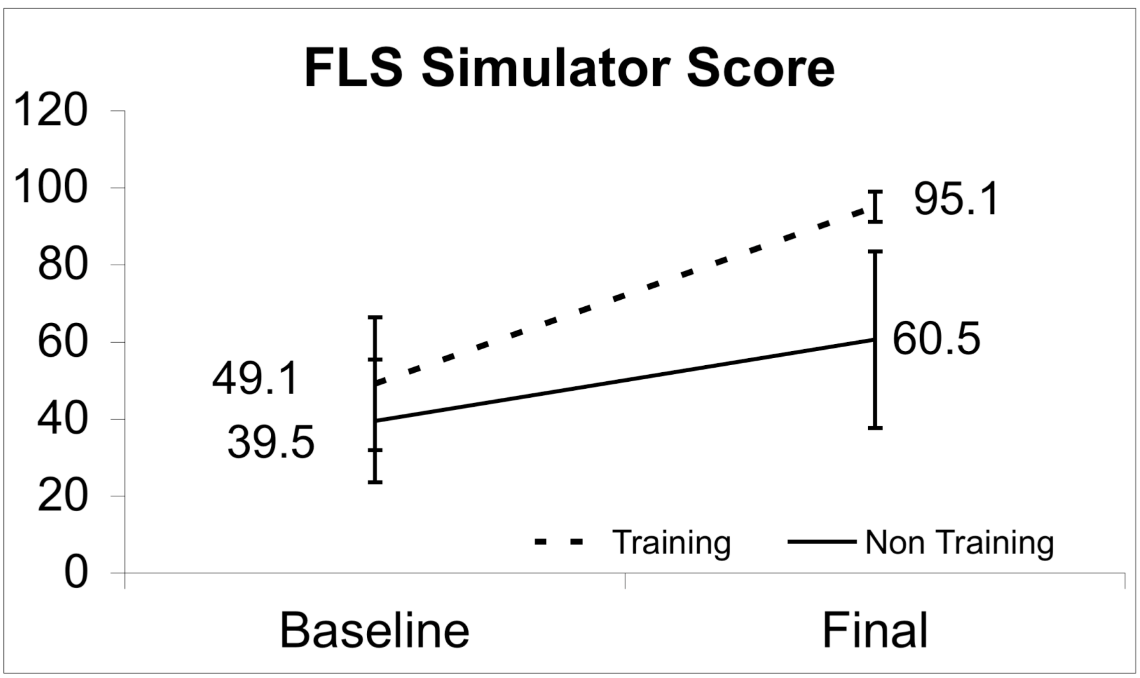


Figure 3: Evaluation of laparoscopic operating room performance during dissection of the gallbladder from the liver bed during laparoscopic cholecystectomy. There was no difference in the groups at baseline (p=0.47). The group trained in the simulator improved more than the non-trained group (p=0.0003).

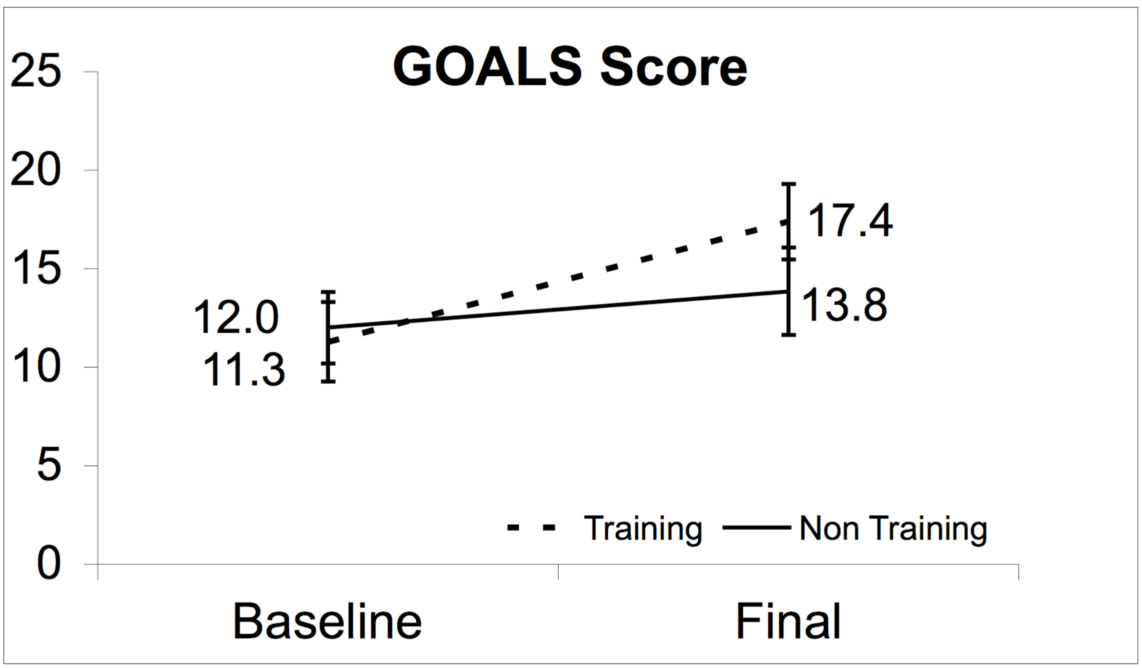
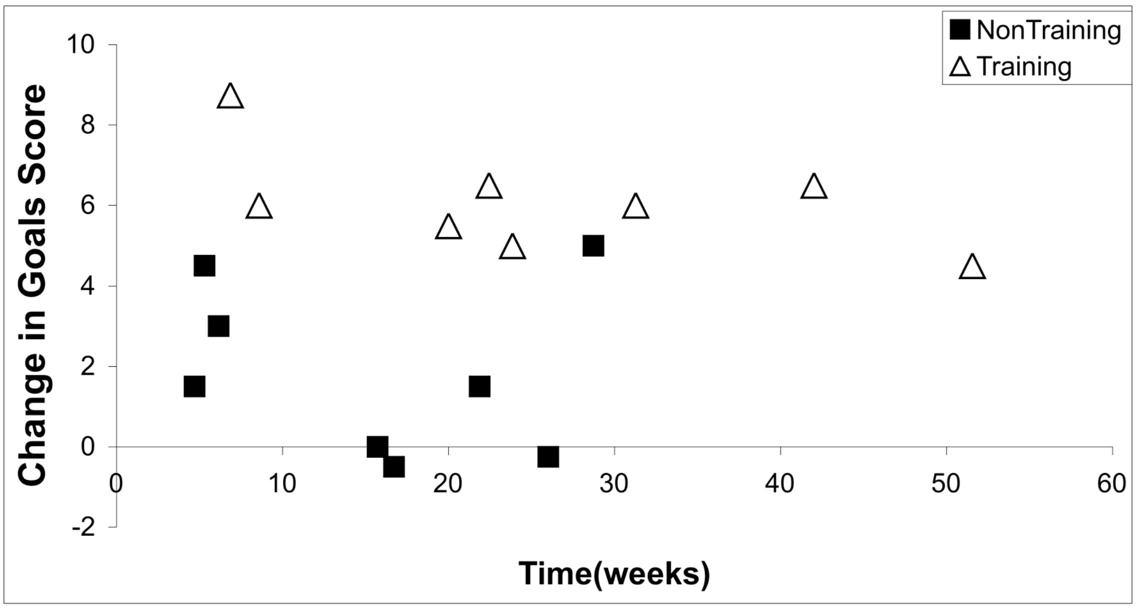


Figure 4: correlation between the time between baseline and final evaluation (x axis) and the improvement in GOALS score (y axis).
Pearson correlation coefficient = 0.13



SECTION F – REFFERENCES

1. Reynolds W, Jr. The first laparoscopic cholecystectomy. *JSLS* 2001; 5(1):89-94.
2. Ellison EC, Carey LC. Lessons learned from the evolution of the laparoscopic revolution. *Surg Clin North Am* 2008; 88(5):927-41, v.
3. Fried GM. Lessons from the surgical experience with simulators: incorporation into training and utilization in determining competency. *Gastrointest Endosc Clin N Am* 2006; 16(3):425-34.
4. McPartland KJ, Pomposelli JJ. Iatrogenic biliary injuries: classification, identification, and management. *Surg Clin North Am* 2008; 88(6):1329-43.
5. Laparoscopic cholecystectomy. *JAMA* 1991; 265(12):1585-7.
6. Marescaux J, Clement JM, Tassetti V, et al. Virtual reality applied to hepatic surgery simulation: the next revolution. *Ann Surg* 1998; 228(5):627-34.
7. Issenberg SB, McGaghie WC, Hart IR, et al. Simulation technology for health care professional skills training and assessment. *JAMA* 1999; 282(9):861-6.
8. Gorman PJ, Meier AH, Rawn C, et al. The future of medical education is no longer blood and guts, it is bits and bytes. *Am J Surg* 2000; 180(5):353-6.
9. Fried GM, Derossis AM, Bothwell J, et al. Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. *Surg Endosc* 1999; 13(11):1077-81; discussion 1082.
10. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. *Ann Surg* 2004; 240(3):518-25; discussion 525-8.
11. Peters JH, Fried GM, Swanstrom LL, 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.

12. http://home.absurgery.org/default.jsp?certgsqe accessed March 28 2009.
13. http://www.sdc.qld.edu.au/fls.htm. *Accessed June 12 2009.*
14. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; 84(2):273-8.
15. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 2005; 190(1):107-13.
16. McCluney AL, Vassiliou MC, Kaneva PA, et al. FLS simulator performance predicts intraoperative laparoscopic skill. *Surg Endosc* 2007; 21(11):1991-5.
17. Satava RM, Cuschieri A, Hamdorf J. Metrics for objective Assessment. *Surg Endosc* 2003; 17(2):220-6.
18. Swanstrom LL, Fried GM, Hoffman KI, et al. Beta test results of a new system assessing competence in laparoscopic surgery. *J Am Coll Surg* 2006; 202(1):62-9.
19. Feldman LS, Hagarty SE, Ghitulescu G, et al. Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. *J Am Coll Surg* 2004; 198(1):105-10.
20. Vassiliou MC, Ghitulescu GA, Feldman LS, et al. The MISTELS program to measure technical skill in laparoscopic surgery : evidence for reliability. *Surg Endosc* 2006; 20(5):744-7.
21. McCloy R, Stone R. Science, medicine, and the future. Virtual reality in surgery. *BMJ* 2001; 323(7318):912-5.
22. Taffinder N, Sutton C, Fishwick RJ, et al. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform* 1998; 50:124-30.

23. Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. *Surg Endosc* 2002; 16(12):1746-52.
24. Botden SM, Buzink SN, Schijven MP, et al. Augmented versus virtual reality laparoscopic simulation: what is the difference? A comparison of the ProMIS augmented reality laparoscopic simulator versus LapSim virtual reality laparoscopic simulator. *World J Surg* 2007; 31(4):764-72.
25. Datta V, Chang A, Mackay S, et al. The relationship between motion analysis and surgical technical assessments. *Am J Surg* 2002; 184(1):70-3.
26. Datta V, Mackay S, Mandalia M, et al. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg* 2001; 193(5):479-85.
27. Smith SG, Torkington J, Brown TJ, et al. Motion analysis. *Surg Endosc* 2002; 16(4):640-5.
28. Francis NK, Hanna GB, Cuschieri A. Reliability of the Advanced Dundee Endoscopic Psychomotor Tester for bimanual tasks. *Arch Surg* 2001; 136(1):40-3.
29. Francis NK, Hanna GB, Cuschieri A. The performance of master surgeons on the Advanced Dundee Endoscopic Psychomotor Tester: contrast validity study. *Arch Surg* 2002; 137(7):841-4.
30. Fried GM, Feldman LS. Objective assessment of technical performance. *World J Surg* 2008; 32(2):156-60.
31. Moorthy K, Munz Y, Sarker SK, et al. Objective assessment of technical skills in surgery. *BMJ* 2003; 327(7422):1032-7.
32. Szalay D, MacRae H, Regehr G, et al. Using operative outcome to assess technical skill. *Am J Surg* 2000; 180(3):234-7.
33. Hanna GB, Frank TG, Cuschieri A. Objective assessment of endoscopic knot quality. *Am J Surg* 1997; 174(4):410-3.

34. Kohn LT, Corrigan jm, Donaldson MS. To err is human: building a safer health system. *National Academy Press* 1999; Washington, DC.
35. Satava RM. The nature of surgical error: a cautionary tale and a call to reason. *Surgical Endoscopy* 2005; 19(8):1014-6.
36. Seymour NE, Gallagher AG, Roman SA, et al. Analysis of errors in laparoscopic surgical procedures. *Surg Endosc* 2004; 18(4):592-5.
37. Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; 73(9):993-7.
38. Dath D, Regehr G, Birch D, et al. Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. *Surg Endosc* 2004; 18(12):1800-4.
39. Scott DJ, Rege RV, Bergen PC, et al. Measuring operative performance after laparoscopic skills training: edited videotape versus direct observation. *J Laparoendosc Adv Surg Tech A* 2000; 10(4):183-90.
40. Vickers AJ, Savage CJ, Hruza M, et al. The surgical learning curve for laparoscopic radical prostatectomy: a retrospective cohort study. *Lancet Oncol* 2009; 10(5):475-80.
41. Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. *J Am Coll Surg* 2007; 204(2):308-13.
42. Ghaderi I, Feldman LS, Sroka G, et al. Evaluation of Surgical Performance During Laparoscopic Incisional Hernia Repair - Preliminary Data. *Canadien Association of General Surgeons Forum* 2009; Victoria BC(Poster Presentation).
43. Vassiliou MC, Feldman LS, Fraser SA, et al. Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. *Surg Innov* 2007; 14(3):211-6.

44. Way LW, Stewart L, Gantert W, et al. Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. *Annals of Surgery* 2003; 237(4):460-9.
45. Derossis AM, Bothwell J, Sigman HH, et al. The effect of practice on performance in a laparoscopic simulator. *Surg Endosc* 1998; 12(9):1117-20.
46. Scott DJ, Bergen PC, Rege RV, et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? *J Am Coll Surg* 2000; 191(3):272-83.
47. Hyltander A, Liljegren E, Rhodin PH, et al. The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surg Endosc* 2002; 16(9):1324-8.
48. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 2002; 236(4):458-63; discussion 463-4.
49. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg* 2004; 91(2):146-50.
50. Korndorffer JR, Jr., Scott DJ, Sierra R, et al. Developing and testing competency levels for laparoscopic skills training. *Archives of Surgery* 2005; 140(1):80-4.
51. Korndorffer JR, Jr., Hayes DJ, Dunne JB, et al. Development and transferability of a cost-effective laparoscopic camera navigation simulator. *Surgical Endoscopy* 2005; 19(2):161-7.
52. Schijven MP, Jakimowicz JJ, Broeders IA, et al. The Eindhoven laparoscopic cholecystectomy training course--improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum. *Surg Endosc* 2005; 19(9):1220-6.

53. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *American Journal of Surgery* 2007; 193(6):797-804.
54. Aggarwal R, Ward J, Balasundaram I, et al. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg* 2007; 246(5):771-9.
55. Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann Surg* 2006; 244(2):310-4.
56. Dosis A, Aggarwal R, Bello F, et al. Synchronized video and motion analysis for the assessment of procedures in the operating theater. *Archives of Surgery* 2005; 140(3):293-9.
57. Sturm LP, Windsor JA, Cosman PH, et al. A systematic review of skills transfer after surgical simulation training. *Ann Surg* 2008; 248(2):166-79.
58. Scott DJ, Bergen PC, Euhus DM, et al. Intense laparoscopic skills training improves operative performance of surgery residents. Vol. 50: Am Coll Surg Surg Forum, 1999. pp. 670-670.
59. Gurusamy K, Aggarwal R, Palanivelu L, et al. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery *British Journal of Surgery* 2008; 95:1088-1097.
60. Madan AK, Frantzides CT. Prospective randomized controlled trial of laparoscopic trainers for basic laparoscopic skills acquisition. *Surgical Endoscopy* 2007; 21(2):209-13.
61. Hamilton EC, Scott DJ, Fleming JB, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surgical Endoscopy* 2002; 16(3):406-11.

62. Munz Y, Kumar BD, Moorthy K, et al. Laparoscopic virtual reality and box trainers: is one superior to the other? *Surgical Endoscopy* 2004; 18(3):485-94.

63. Matsumoto ED, Hamstra SJ, Radomski SB, et al. The effect of bench model fidelity on endourological skills: a randomized controlled study. *Journal of Urology* 2002; 167(3):1243-7.

64. Grober ED, Hamstra SJ, Wanzel KR, 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.

65. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005; 241(2):364-72.

66. Ritter EM, Scott DJ. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. *Surg Innov* 2007; 14(2):107-12.

67. Verdaasdonk EG, Dankelman J, Lange JF, et al. Incorporation of proficiency criteria for basic laparoscopic skills training: how does it work? *Surgical Endoscopy* 2008; 22(12):2609-15.

68. Fraser SA, Klassen DR, Feldman LS, et al. Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. *Surg Endosc* 2003; 17(6):964-7.

69. Grantcharov TP, Schulze S, Kristiansen VB. The impact of objective assessment and constructive feedback on improvement of laparoscopic performance in the operating room. *Surg Endosc* 2007; 21(12):2240-3.

70. Derossis AM, Fried GM, Abrahamowicz M, et al. Development of a model for training and evaluation of laparoscopic skills. *Am J Surg* 1998; 175(6):482-7.

71. Soper NJ, Fried GM. The Fundamentals of Laparoscopic Surgery: Its time has come. *Bull Am Coll Surg* 2008; 93(9):30-32.

72. Korndorffer JR, Jr., Dunne JB, Sierra R, et al. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *J Am Coll Surg* 2005; 201(1):23-9.

73. http://www.flsprogram.org accessed March 28 2009.

74. Scott DJ, Ritter EM, Tesfay ST, et al. Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training. *Surg Endosc* 2008; 22(8):1887-93.

75. Satava RM. Historical review of surgical simulation--a personal perspective. *World J Surg* 2008; 32(2):141-8.

76. Dauster B, Steinberg AP, Vassiliou MC, et al. Validity of the MISTELS simulator for laparoscopy training in urology. *J Endourol* 2005; 19(5):541-5.

77. Healy GB. The college should be instrumental in adapting simulators to education. *Bulletin of the American College of Surgeons* 2002; 87(11):10-1.

78. Damiani G, Pinnarelli L, Sammarco A, et al. Postoperative pulmonary function in open versus laparoscopic cholecystectomy: a meta-analysis of the Tiffenau index. *Digestive Surgery* 2008; 25(1):1-7.

79. Buunen M, Gholghesaei M, Veldkamp R, et al. Stress response to laparoscopic surgery: a review. *Surgical Endoscopy* 2004; 18(7):1022-8.

80. Haq Z, Rahman M, Siddique MA, et al. Interleukin-6 (IL-6) and tumour necrosis factor-alpha in open and laparoscopic cholecystectomy. *Mymensingh Med J* 2004; 13(2):153-6.

81. Ellison EC, Carey LC. Lessons learned from the evolution of the laparoscopic revolution. *Surgical Clinics of North America* 2008; 88(5):927-41, v.

82. Lucas SM, Zeltser IS, Bensalah K, et al. Training on a virtual reality laparoscopic simulator improves performance of an unfamiliar live

laparoscopic procedure. *Journal of Urology* 2008; 180(6):2588-91; discussion 2591.

83. Hsu KE, Man FY, Gizicki RA, et al. Experienced surgeons can do more than one thing at a time: effect of distraction on performance of a simple laparoscopic and cognitive task by experienced and novice surgeons. *Surg Endosc* 2008; 22(1):196-201.

Acknowledgments

I want to express my gratitude to Dr Liane Feldman, my supervisor, and to Dr Gerald Fried, head of the Steinberg-Bernstein Center for Minimally Invasive Surgery, for the opportunity to take a part in such an important educational project, and for their unlimited support and guidance.

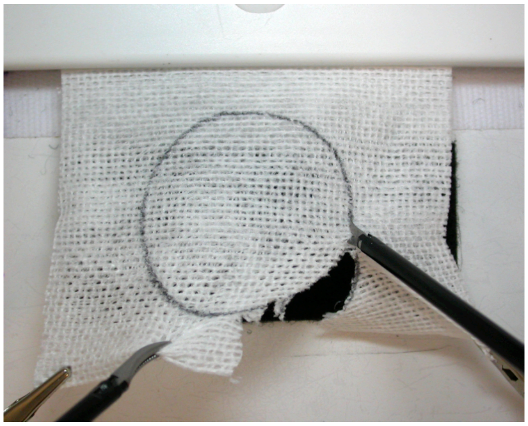
I thank Dr Melina Vassiliou and Dr Raad Fayez for their important assistance, and Pepa Kaneva for the long hours we spent analyzing the data.

Last but not least – I thank Inbal, my wife, who's sacrifice made all this possible.

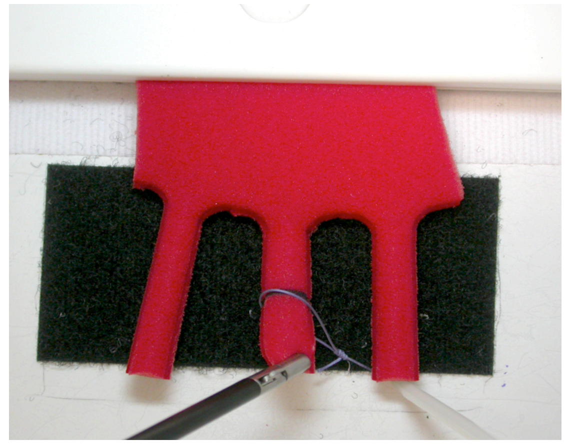
This work could never have been carried out without the structure and resources made available to me by the Steinberg-Bernstein Center for Minimally Invasive Surgery. This research has been supported by an unrestricted educational grant from Covidien Canada. I also received a Fellowship Grant from the American Physician Fellowship for medicine in Israel.

Appendix I: The FLS Simulator and Tasks

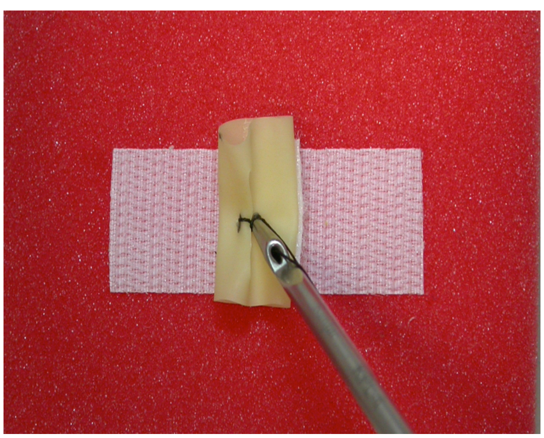
1-a: The FLS Simulator with task no.1: Peg Transfer



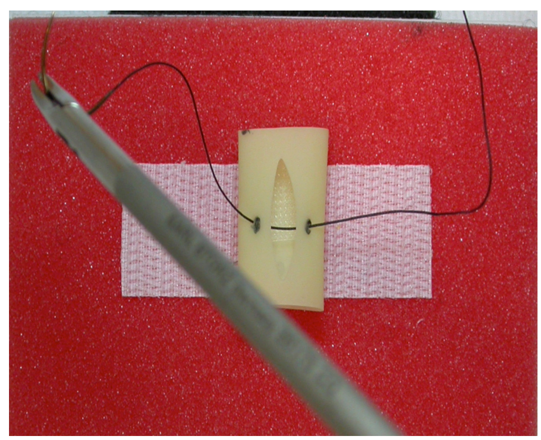
1-b: task no.2: Circle Cut



1-c: task no.3: Endo-Loop Placement



1-d: task no.4: Extra-corporeal Stitch



1-e: task no.5: Intra-corporeal Stitch

Appendix II – FLS Performance Metrics
FLS Score Sheet

Candidate: Evaluator: Date:

Trial number:

	Raw score: time - penalty	Normalized score
PEG TRANSFER		raw/**237** x 100 = x .42
Time: 300 secs – time = (in seconds)		
Penalty: (Pegs not transferred/6)*100 =		
PATTERN CUTTING		raw/**280** x100 = x .36
Time: 300 secs – time =		
Penalty: (area * 100)/33 =		
ENDOLOOP		raw/**142** x 100 = x .70
Time: 180 secs – time =		
Penalty: #mms. away from all pre-drawn lines (1/mm) 50 points for insecure knot		
EXTRACORPOREAL KNOT		raw/**297** x 100 = x .34
Time: 420 secs – time =		
Penalty: # mm. from pre-drawn dots, # mm. gap in incision; Slipping knot = 10; Knot comes apart = 20; Drain avulsion = score of 0		
INTRACORPOREAL KNOT		raw/**520** x 100 = x .19
Time: 600 secs – time =		
Penalty: # mm. from pre-drawn dots, # mm. gap in incision; Slipping knot = 10; Knot comes apart = 20; Drain avulsion = score of 0		
Total score		Final Normalized Score:
Score/5 = total normalized score		

Appendix III - GOALS

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: