Design, validation and implementation of a virtual reality high fidelity laparoscopic appendicectomy curriculum

Daniel Moshe Sinitsky

MD (Res)

UCL

November 2021

Declaration

I, Daniel Moshe Sinitsky confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Daniel M Sinitsky
Date: 9th November 2021

Abstract

Introduction

The treatment for acute appendicitis is laparoscopic appendicectomy (LA), usually performed by trainees who face significant challenges to training. Simulation curricula are being increasingly utilised and optimised to accelerate learning and improve skill retention in a safe environment. The aim of this study is to produce and implement a virtual reality (VR) curriculum for laparoscopic appendicectomy (LA) on the high-fidelity LAP Mentor VR simulator.

Methodology

Performance data of randomised experts and novices were compared to assess the construct validity of the LAP Mentor basic skills (BS) and LA modules. Face validity of the simulator and module was assessed by questionnaire. These results informed the construction of a VR LA curriculum on an evidence-based theoretical framework. The curriculum was implemented and evaluated by analysis of participant diaries.

Results

Thirty-five novices and 25 experienced surgeons performed either BS, five LA procedural tasks or the LA full procedure. Both modules demonstrated construct validity. The LA module was deemed moderately realistic and useful for developing laparoscopic psychomotor skills. Seven novice trainees completed the new LA curriculum (three others dropped out). Analysis of participants diaries revealed the presence of frustration, the benefits of feedback sessions and the advantages and pitfalls of open access.

Discussion

Evaluations of the implementation of similar curricula are rare and participant diaries led to critical insights. The curriculum was difficult and sometimes frustrating, mitigated by rewarding experiences and coaching. The latter facilitated deliberate practice. Scheduling

issues were mitigated by open access. Limitations of the curricula include the invariability in the presentation of appendicitis, and the reason for dropouts are not known.

Conclusion

Several BS and all LA tasks are construct-valid. A new VR LA curriculum was implemented and analysis of participant diaries yielded critical insights into real-world implementation. Future study should investigate its effect on real-world performance and patient outcomes.

Impact Statement

Surgical training in the United Kingdom is undoubtedly facing severe challenges as a result of a reduction in working hours. This has adversely affected the training opportunities in the operating room and has led to training discontinuity and fragmentation due to changing working patterns. Meanwhile, laparoscopic surgery has become ubiquitous in General Surgery, yet the required skills are largely not transferable from open surgery. Therefore, in order to accelerate learning while also maintaining patient safety, it has become imperative for junior surgical trainees to learn the necessary laparoscopic skills in a safe simulated environment. This is particularly true for laparoscopic appendicectomy, since it is a very common procedure that is mainly performed by trainees.

Up until now, curricula have been developed around simulators for virtual reality (VR) laparoscopic surgery. The quality of these curricula has been variable, often only loosely bound to valid theoretical frameworks that attempt to optimise the acquisition and retention of laparoscopic skills. As this thesis outlines, there are very many aspects to consider when designing a curriculum. For example, the designation of spacing intervals between practice sessions or how intrinsic design can affect trainee motivation.

The Lap Mentor is a high-fidelity VR laparoscopy simulator, upon which there exists basic skills and appendicectomy modules. There are very few studies that have validated and developed a curriculum based upon similar modules, and even rarer are the published experiences of real-world implementation. The present study is the first to have developed *and* implemented a curriculum in the real-world. It may also be the most robust in terms of its attention to up-to-date evidence regarding curriculum development and its foundation upon a strong theoretical framework. Indeed, the validation and development phase of the present study has subsequently been published in The American Journal of Surgery.* This was also orally presented at national and international conferences.**

The latter phase of the present study is a rare evaluation of real-world implementation, which uniquely utilises a qualitative analysis of participant diaries. This has yielded critical insights into poorly understood facets of curriculum design and implementation, particularly regarding the role of human instruction and feedback. It is anticipated that this phase will also be published in a peer-reviewed journal and that it is likely to positively influence the simulation curricula of other institutions around the world.

The real-world outcomes of training using this new curriculum are yet to be tested (such as surgical performance, patient outcomes and the effect on surgical training more broadly). However, the existing literature has already demonstrated translation of skills from the VR environment to real-world surgical performance, and so it is anticipated that positive effects are already being felt by those junior trainees who have already completed the curriculum and, most importantly, by their patients.

** Sinitsky DM, Fernando B, Potts H, et al. Development of a structured virtual reality curriculum for laparoscopic appendicectomy. Am J Surg, 2020;219(4):613-621.*

*** Sinitsky D, Fernando B, Potts H, Lykoudis P, Hamilton G, Berlingieri P (presented by Sinitsky D). Development of a structured virtual reality curriculum for laparoscopic appendectomy. International presentation (selected abstract). 25th Annual Meeting of the Society in Europe for Simulation Applied to Medicine (SESAM)*
Session: Curriculum development, Society in Europe for Simulation Applied to Medicine (SESAM), Audience: Conference delegates, 13th June 2019, George Moore M137, Glasgow Caledonian University, Glasgow, United Kingdom.

Sinitsky D, Fernando B, Potts H, Lykoudis P, Hamilton G, Berlingieri P (presented by Sinitsky D). Development of a structured virtual reality curriculum for laparoscopic appendectomy: widening the accessibility to high-fidelity simulation technology for surgical trainees. Local presentation (selected abstract), UCL Education Conference, Session: Digital education and innovations, Chair: Dr Mina Sotirou, Audience: Conference delegates, 1st April 2019, Room 744, UCL Institute of Education (IOE), London, United Kingdom.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 General Surgery

General Surgery is a broad term that describes an evolving specialist field of surgery, encompassing the subspecialties of upper gastrointestinal (GI), colorectal, hepatopancreaticobiliary, breast and endocrine surgery. The term reflects an era that had seen surgeons routinely operate across all subspecialties. Nowadays, the elective (non-emergency) caseload is often managed by general surgeons who have the relevant subspecialty expertise (for example, oesophageal cancer by upper GI surgeons, complicated diverticular disease by colorectal surgeons), though with some overlap with conditions such as abdominal wall hernia. As knowledge and experience has increased, and as the evidence suggests that high operative volume is associated with better patient outcomes (1), General Surgery is gradually becoming devolved into its constituent subspecialties. Vascular surgery is an example, having been recently divorced from the General Surgery training programme by the Joint Committee on Surgical Training (JCST) (2).

Despite this trend, the general surgeon still requires cross-subspecialty flexibility in the emergency setting. She or he is routinely required to diagnose and manage emergency conditions of other subspecialties, but which still fall within the remit of General Surgery. Indeed, the Intercollegiate Surgical Curriculum Programme (ISCP) syllabus for the United Kingdom's General Surgery training programme clearly outline the emergency conditions for which all general surgeons must be able to demonstrate competence (3). For example, a perforated peptic ulcer frequently requires emergency surgery to the stomach or duodenum, the traditional battleground of the upper GI surgeon, and acute appendicitis is, strictly speaking, a colorectal condition.

Yet acute appendicitis is the most common abdominal emergency requiring emergency surgery (4), and traditionally among the first laparoscopic procedures performed by trainees.

1.2 Surgical Training in the United Kingdom

Over the past few years, surgical training in the United Kingdom has undergone significant changes. Without adapting to these new challenges and pressures, the quality of tomorrow's surgeons is threatened.

Before the Modernising Medical Careers (MMC) reforms in 2007, junior doctors wishing to pursue a career in surgery would research consultant trainers around the country and apply for posts within those firms. Each consultant would themselves screen resumes and interview prospective trainees before offering a post to he or she who was deemed most suitable. Trainees could do this repeatedly as a Senior House Officer (SHO) or Specialist Registrar (SpR) until they felt ready to progress on the path to becoming a fully-trained surgeon.

Applications for surgical training are now anonymised and submitted to a central processing team and there is only a single interview for each stage of training: Core Surgical Training years CT1 and CT2 (CT, Core Trainee), and Specialty Registrar (StR) years ST3 to ST8 (for General Surgery). The nature of this system encourages the 'run-through' of trainees, meaning that they are no longer able to easily add posts and years to their training time as desired or required, but more likely would simply fail to progress, or more worryingly, become consultants without the breadth and depth of experience of their predecessors. Indeed, and with great controversy, MMC was set to shorten training time to produce more consultants, albeit with less experience (5).

Prior to MMC, SHOs were often in posts for between six to 12 months at a time. However, the modern CT1 year is typically made up of three four-month posts, followed by two six-month posts during CT2. Trainees may therefore find it difficult to build a relationship and trust with their trainer within such a short space of time, sufficient to afford them the operating experience required to progress within the operating room (OR). This may be one of many reasons why MMC has led to the belief that the quality of surgical training is set to fall (6).

These challenges to surgical training have been compounded by the European Working Time Directive (EWTD) that limits working time to 48 hours per week. Whilst this may not have reduced patient safety (7), and may have decreased the incidence of medical error (8), the total time available to train a surgeon has drastically reduced.

These working time restrictions have necessitated changes to working patterns in order to maintain service delivery – they may have actually led to trainee doctors working *more* unsocial hours than in the past (9). The negative effect of this change in working patterns on training is four-fold:

1. Out-of-hours shifts are typically less supervised than during normal working hours;
2. In order to improve patient safety, the National Confidential Enquiry into Patient Outcomes and Deaths (NCEPOD) has resulted in a great reduction in the opportunities for training in emergency surgery out-of-hours (10);
3. The increase in out-of-hour shifts results in mandatory ‘rest’ periods that fall during daylight hours, where training opportunities for complex elective surgery are missed;
4. The lack of care continuity can result in a profound reduction in the number of ‘take-back’ cases – excellent learning opportunities are missed by not being present to recognise and manage complications from their own procedures (11).

Accordingly, of 2056 medical graduates of 2002 surveyed in 2013-2014, 85% of those in surgical specialties felt that the EWTD had indeed had a negative impact on their training opportunities (12).

Ultimately, new holders of the Certificate of Completion of Training (CCT) in surgery must be as competent as ever, in the face of severe challenges. As such, it is vital to explore all means and methods to make the most of the available time in training.

Despite narrowing of the window for surgical training, the volume of necessary skill that must pass through it has arguably increased. This is particularly demonstrable given the widespread adoption of laparoscopic surgery over the past 20 years.

1.3 Laparoscopic Surgery

Laparoscopy is the term that describes the use of ‘keyholes’ to operate with long instruments inside the abdomen. Compared to traditional open surgery, laparoscopy tends to confer a reduction in pain and wound complications, shorter hospital stay and quicker recovery (13). While working hours have shortened, the utilisation of laparoscopy has increased. This is relevant – while some skills are transferable between laparoscopic and open surgery (such as the recognition of tissue planes, effective tissue retraction and general dissection methods), many psychomotor skills required for laparoscopy are unique, they are not transferable from the open operating environment, and they need to be learned from the start.

There are numerous psychomotor challenges inherent in laparoscopic surgery. It involves the use of long instruments with inverted movements (due to the *fulcrum* effect of a pivoting instrument) while observing a two-dimensional (2D) projection elsewhere. The unique psychomotor challenges have been neatly summarised by Munz *et al* (14):

- Shift from three-dimensional (natural) view to 2D monitor display;
- Impaired judgement of depth perception and spatial relationships;
- Video-eye-hand coordination;
- Adaptation to fulcrum effect;
- Manipulation of long instruments while adjusting for amplified tremor and fewer degrees of freedom;
- Reduced haptics (‘force feedback’) perception.

It is therefore of little surprise that the early part of the learning curve, defined as “the number of operations to reach an experience level with a low complication rate” (15,16), is associated with an increased rate of surgical complications. This association was borne out of analyses of laparoscopic cholecystectomies during laparoscopy’s formative years. An often-cited prospective multi-centre study of 1,518 patients undergoing laparoscopic cholecystectomy revealed that the rate of bile duct injury within the first 13 laparoscopic cholecystectomies was 2.2%, compared to 0.1% for subsequent patients (17). Similarly, in a retrospective survey-based analysis of greater than 77,000 patients who underwent laparoscopic cholecystectomy throughout over 4,000 hospitals in the United States (US), the rate of bile duct injury was significantly higher for those surgeons who had performed less than 100 procedures (18).

The suggestion of a learning curve in these analyses is surprising given that these surgeons are highly likely to have reached expert level with traditional open cholecystectomy long before attempting the laparoscopic equivalent. This suggests the non-transferability of open surgical skill to laparoscopy, with a new learning curve and its associated complications.

Traditionally, learning an operation is a process that sees the trainee take on gradually more complex and risky parts of a procedure under ever-decreasing levels of supervision, which is an example of learning by “chunking” (19). However, in addition to the unique psychomotor challenges of laparoscopy compared to open surgery, there are also important unique challenges of supervision inherent in teaching laparoscopic surgery in the OR that may compound the effect of the learning curve on patient safety. For example, it is possible for both supervisor and trainee to have their hands technically engaged in the same aspect of an open procedure, whereas this is often impossible during laparoscopic surgery. Furthermore, during open surgery, resuming technical control of an open operation is often no more than a few degrees of movement between the hands of supervisor and trainer, and can often occur instantly. However, during laparoscopic surgery, the trainee is first required to stop before stepping aside and handing over the instruments. These challenges of both psychomotor skills and supervision add to a sense of anxiety for the responsible clinician who is observing a trainee ascend his or her learning curve on real patients.

Despite all these barriers, it may be of some surprise to learn that laparoscopic appendicectomy (LA), the treatment of choice for acute appendicitis, is an extremely common procedure that is most often performed by trainees (4) on a cohort of patients that are, by definition, acutely unwell.

1.4 Acute Appendicitis & Surgical Training

Acute appendicitis is the most common abdominal surgical emergency (20), accounting for nearly 48,000 admissions in the UK during 2018-2019 (21). It is an acute inflammatory condition of the vermiform appendix of uncertain aetiology. The gold standard of treatment is appendicectomy (surgical excision of the appendix), although there is a place for antibiotic therapy, for example in those who decline or are deemed unsuitable for surgery (22).

Traditionally, open appendicectomy is performed via a grid-iron (*McBurney*) incision in the right iliac fossa as the standard operative approach. Since Semm's description of laparoscopic appendicectomy in 1983 (23), this minimally-invasive technique has become much more common, now accounting for two-thirds of all appendicectomies in the United Kingdom (4). The main advantages of laparoscopic over open appendicectomy include cosmesis, shorter hospital stay, less pain and a reduced wound infection rate, albeit at a cost of a higher risk of intra-abdominal abscess (24).

As a consequence of diagnostic uncertainty, many patients with suspected appendicitis undergo appendicectomy where the appendix appears normal. As such, 12% of males and 23% of females will undergo appendicectomy during their lifetime (20). According to a prospective study of 3,326 patients across 95 centres within the United Kingdom, 90% of these were performed by trainees (49% were those at or below the level of Specialist Trainee Year 5, ST5), of which only 24% were supervised with a consultant present in the OR (4). Therefore, it is imperative that the trainee has quickly learned the skills necessary to perform the task safely. As succinctly put by Larsen *et al* – "it is an ethical imperative that an operation performed by a supervised novice ought to have the same outcome as that of the supervisor" (25).

1.5 Laparoscopic Appendicectomy – Technique

Prior to LA, the trainee must have learned to understand the selection criteria for laparoscopic surgery in those patients who are suspected of having acute appendicitis. For example, patients may be too small (infants and children) or too large (obese adults), or may have concomitant cardiorespiratory problems that would be exacerbated by abdominal insufflation pressures during laparoscopy. Another common relative contra-indication is a history of previous abdominal surgery, or other reasons that would raise the likelihood of peritoneal adhesions. Poor patient choice can easily lead to serious morbidity and even mortality, and so it is vital that there is a proper understanding of this crucial decision-making process.

Once the decision has been made with the patient, informed consent has been obtained and the patient has been anaesthetised, LA can be divided into four key stages. These are Position & Preparation, Abdominal Access, Appendicectomy and Closure, each of which includes a series of key steps. Although there is some variation in practice, in general the patient undergoing LA is carefully placed in a supine position, and appropriately secured to the table. The World Health Organization (WHO) checklist is followed, which includes steps such as hair removal and antibiotic administration (26). Skin preparation and draping is performed, and the laparoscopic equipment is set up in a manner that affords the surgeon maximal comfort during the procedure. Access is obtained either by blunt, open or Veres needle methods, avoiding bowel or vascular injury, and an appropriate insufflation pressure is introduced and maintained. Usually, two further ports are positioned under direct vision. The appendix is identified, mobilised and the mesoappendix is detached while maintaining haemostasis. The base of the appendix is then ligated and the appendix is delivered. Peritoneal lavage is carried out as deemed appropriate, a final check is performed before removing the ports, taking care to identify and manage any port-site bleeding prior to closure. Finally, fascial and skin closure are performed.

Each of these steps confers specific risks of which the surgeon must be aware, and the trainee must learn to minimise. These are outlined in Table 1.

Key Step	**Associated Risks**
Position & Preparation	
Patient positioning (supine)	Patient may slide from table
Bladder decompression*	Predisposition to bladder injury
Skin preparation	Surgical site infection
Draping of operative field	Uncomfortable operating position
Laparoscopic equipment positioning	
Abdominal access	
Creation of pneumoperitoneum	Visceral or vascular injury, carbon dioxide (CO_2) embolisation, cardiac failure, cardiac arrest
Place two further ports	Visceral or vascular injury
Appendicectomy	
Identification of appendix	Resection of wrong organ/structure, incomplete excision
Mobilisation of appendix	Visceral injury (caecum, ascending colon, small bowel)
Detachment of mesoappendix	Intra-/post-operative bleeding
Ligation & resection of appendix	Stump leak leading to sepsis/colocutaneous fistula
Delivery of appendix	Retention of appendix
Wash as necessary*	Abscess formation
Final inspection	Missed injury
Closure	
Port removal	Post-operative bleeding (inferior epigastric vessels)
Fascial closure	Small bowel injury leading to sepsis/fistula
Skin closure & dressing	Poor cosmesis, retained suture material

Table 1 – Key Steps & Intraoperative Risks of Laparoscopic Appendicectomy with corresponding risks. **Denotes aspects where controversy might exist and so may be discretionary.*

Minimisation of risk by a surgeon during any procedure can be considered as a product of two components: decision-making and psychomotor skill. For example, the surgeon must decide the safest method for creating a pneumoperitoneum in a patient who has had prior surgery, or whether or not the tissue that he or she is grasping is indeed the appendix. To a large degree, any shortcoming in a trainee's intraoperative decision-making ability can be supplanted by that of the supervisor who needs not touch the patient. However, inadequate psychomotor skill, such as that required in the physical action of dissecting the mesoappendix and ligating the appendix base, can only be overcome by the supervisor physically taking over the procedure. It follows, therefore, that a trainee is much more likely to progress if he or she arrives in the operating theatre already at the required standard of psychomotor skill.

1.6 Learning Laparoscopic Psychomotor Skills Outside of the Operating Environment

Learning laparoscopic psychomotor skills outside of the operating environment is not a new concept. Indeed, there are already courses available that aim to teach these skills to trainees. Two such courses held at the Royal College of Surgeons of England are the 2-day Core Skills in Laparoscopic Surgery and 2-day Intermediate Skills in Laparoscopic Surgery courses, priced at £1,099 and £666.00 per participant, respectively (27,28). Neither of these are mandatory. Their learning outcomes are outlined in Table 2.

RCSE Learning Outcomes

Core Skills in Laparoscopic Surgery

> Identify techniques for access and closure
> Demonstrate a range of suturing and knot-tying techniques
> Perform key laparoscopic surgery procedures with improved hand–eye coordination, depth perception and surgical precision
> Demonstrate knowledge of common laparoscopic procedures, including peptic ulcer repair, appendicectomy and cholecystectomy
> Assess and anticipate a range of complications and demonstrate how to deal with them

Intermediate Skills in Laparoscopic Surgery Learning Outcomes

> Outline the principles of good ergonomic port placement and surgeon and instrument positioning
> Demonstrate a range of advanced suturing and knot tying techniques
> Perform advanced laparoscopic surgery procedures including closure of a hiatus hernia and a secure fundoplication on a model
> Demonstrate how to use an ultrasonic dissector effectively
> Demonstrate wide-ranging knowledge in laparoscopic surgery and operative procedures, including gastroenterostomy and laparoscopic pyloroplasty

Table 2 – Learning Outcomes for both Core and Intermediate Skills in Laparoscopic Surgery courses. *These are provided by The Royal College of Surgeons of England (RCS) (27,28).*

The learning outcomes for these courses comprise the acquisition of both knowledge (“demonstrate wide-ranging knowledge”) and psychomotor skill (“perform key laparoscopic surgery procedures”, and “demonstrate…techniques”). Attainment of useful knowledge facilitates safe decision-making. However, much has been learned regarding skill acquisition since the launch of these courses which puts into doubt their utility in achieving this objective. It is vital to understand this evidence-base and how it may be applied in order to build an effective laparoscopic skills curriculum. The key aspects that will be discussed in the design of such a curriculum are:

1. Deliberate practice;
2. Coaching and instruction;
3. Distributed versus massed practice;
4. The role of simulated laparoscopy and virtual reality.

1.6.1 Deliberate Practice

Traditionally, expert performance has been believed to result mainly from innate talent and/or accumulated experience in a specialist field. A surgeon with 20,000 logged operative cases may be expected to be more skilled in surgery than he or she with half that number. Anders Ericsson, the authority on expert performance, challenged this assumption and coined the term ‘deliberate practice’ (DP) to describe the learning behaviours common to those who consistently perform with an extremely high level of skill (29). These are individuals who seek to achieve specific goals that further their level of skill through repetitive and concentrated practice on aspects of their field of expertise (often weaker ones), modifying their techniques according to feedback that they receive. This is different to ‘practice’, which can be defined as merely the engagement of domain-related activities (30).

The requirements, therefore, for DP are as follows:

1. The individual must be highly-motivated (for example, a trainee passionate about surgery);
2. Active conscious engagement in repetitive practice (the trainee fully concentrates and consciously seeks improved performance through each individual repetition);
3. The opportunity and resources for this practice must exist (the trainee has unlimited access to the training environment and necessary equipment);
4. There must be clear, defined goals (new, improved standards are set as the trainee progresses);
5. Provision of feedback (the trainee is receptive to feedback from a coach, or otherwise, to enable appropriate modifications in technique that can affect further improvement in performance).

The learning curves of those who engage in DP plateaus beyond the most traditionally experienced 'experts' who do not engage in it, who instead perform more passively in tasks, almost exclusively in the third, automated, phase of the Fitts & Posner model of learning (31). This latter group suffers what has been termed 'arrested development' (Figure 1), and may be the reason why experience *per se* may not predict performance (32,33). Such DP, often-cited at 10,000 hours-worth, is believed to be responsible for the success of elite performers in fields such as wrestling, karate, chess (34), figure-skating and music (35).

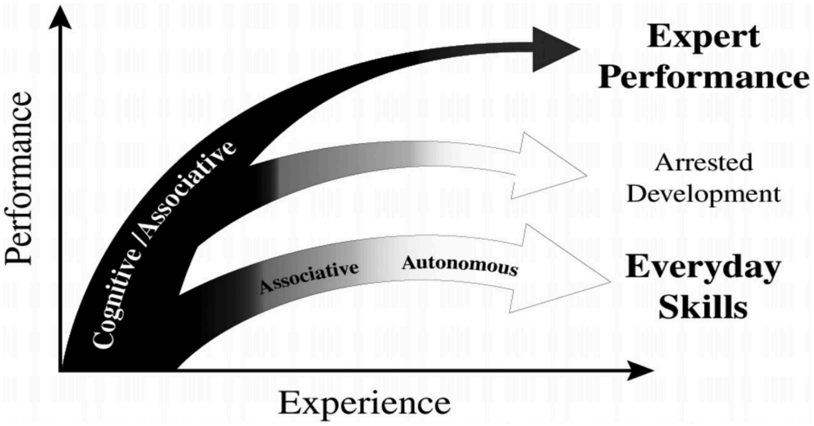


Figure 1 – Deliberate practice and arrested development. *Reproduced from Ericsson* et al *(33). The learning curve for 'everyday skills' reflects that required for tasks to be performed autonomously. The learning curves for 'expert performance' and 'arrested development' reflect the difference between mere practice and deliberate practice.*

By this token, medical and surgical trainees should be learning and practicing their technical skills outside of the OR as a means to further their participation in DP. Just as an elite professional football player engages in DP at the training ground (concentrating on improving penalty kicks or set-pieces with the team, rather than just playing in many games of football), the medical simulation setting is where medical and surgical trainees will find similar opportunities to accelerate learning and improve performance.

To date, there have been a wide range of simulation-based training tools that have a demonstrated ability to accelerate skill acquisition. These include areas such as spinal anaesthesia (36), central venous catheter placement (37), congenital heart surgery (38), coronary artery suturing (39), ultrasound-guided regional anaesthesia (40) and colonoscopy (41), amongst others. In a meta-analysis by McGaghie *et al* comparing traditional medical education with simulation-based medical education (SBME) combined with DP, a large effect size was demonstrated in favour of the latter (42). However, the ubiquitous use of the term 'deliberate practice' across many studies of simulation training appear to attribute to it the beneficial effects of simulator training, when in fact the inadequacy of almost all study control groups renders it impossible to distinguish between the effect of mere practice and that of DP (36,39,43,44).

For example, Nesbitt *et al* trained 10 fourth-year medical students on a porcine simulator to perform coronary artery anastomosis, concluding (as the title of their paper states) that 'medical student deliberate practice can achieve equivalency to senior surgery residents'. Whilst their study is commendable, it would be impossible to deduce whether the medical students' skill acquisition is a result of DP or mere practice alone, due to the absence of a control group. In the following question and answer session, one of the authors also recognised that forbidding practice between simulator sessions isolates the effect of their simulator, but in doing so this limits opportunity that is a pre-requisite for DP (39). In another study by Udani *et al*, 21 anaesthesiology residents were trained to perform subarachnoid spinal anaesthesia on a simulator and randomised to either the intervention or control groups. Both groups completed the 'base curriculum', but the intervention group *also* underwent 'simulator-based deliberate practice' with expert guidance and subsequently performed better on the simulator post-test, stating that 'deliberate practice added a significant, independent, incremental benefit' (36). Though perhaps it is unsurprising that the group who practiced more performed better – like others who cite simulation-based DP, this study actually does very little to examine the effects of the *deliberate* component on learning.

In another example, using the LAP Mentor high-fidelity virtual reality laparoscopy simulator, Hashimoto *et al* randomised 14 residents into DP or control groups. Each participant completed 10 sessions of two simulated laparoscopic cholecystectomies. While the DP group underwent 30 minutes of such practice in each session, focusing on weaker aspects of their performance, the control group was instead assigned 30 minutes of unrelated activity. Post-tests on porcine cadavers demonstrated higher quality of performance in the DP group compared to control (45). Whilst this study suffers the same shortcomings in its control group to isolate the effect of deliberate over mere practice, a hint as to the beneficial effect of DP is evident in the shape of the learning curves – the intervention group's curve plateaued earlier and higher than control, which might suggest that without instructor feedback and/or participants' concentration on the weaker aspects of performance, continued practice by participants in the control group may not have increased their performance due to the demonstration of a plateau, insinuating arrested development.

Addressing the insufficiency of their previous methodology, Udani *et al* returned in 2016 with perhaps the only study published that genuinely compares DP to non-DP simulation-based practice. 28 anaesthesiology residents were randomised to either DP or control group in the acquisition of skill on a simulation of ultrasound-guided regional anaesthesia (UGRA). Both groups received a curriculum that included the UGRA checklist and an exemplar video of the procedure being described and demonstrated. The DP group received coaching, concentrating on parts of the checklist deemed unsatisfactory, until all parts were complete. If further unassisted repetition was deemed satisfactory according to the checklist, then the module was complete. If it was unsatisfactory, then further DP continued. In the control group, rather than undertaking unrelated activity as in previous studies, this time they were instructed to continue with self-directed study/practice as each control participant desired. All participants then underwent a post-test on the simulator both after completion of their training and at a 3-month retention interval. The results were surprising: there was no difference in checklist completion or global performance assessments between groups at both post-tests. Moreover, the control group only practiced for a median of 7 minutes compared to 48 minutes in the DP group (40). It is not immediately clear what is responsible for these findings; it may be related to the level of complexity of the task and the type of skill that is required. It is possible that the assessment of UGRA is largely related to the successful memorisation of a sequence of steps, rather than the psychomotor skill required in actually performing them. In other words, DP and each of its components may confer differential benefits to learning that depend on many factors, not only in the task type and skill required, but also the goals that are set, the presence of a coach, and the type of feedback and the manner in which it is delivered. Unfortunately, much energy has been expended confirming the somewhat obvious effects of *practice* rather than the components that make *deliberate* practice worthwhile.

1.6.2 Coaching & Instruction

So far, despite a plausible rationale, we have not been able to convincingly demonstrate evidence in support of DP over mere practice in the acquisition of clinical skills, probably as a result of poor study design. In any case, it is difficult to imagine how learning could take place without the individual, or any system such as a computer, receiving some form of feedback. He or she who hits tennis balls into the dark will not improve his or her game like the enthusiast who practices in daylight, able to see and therefore appropriately modify the effects of changes in grip, stance and swing according to the trajectory of the ball. Therefore, feedback is a pre-requisite for DP and skill development towards mastery.

The term ‘feedback’ is often confused with ‘instruction’, although the distinction is critical. The latter implies that the information that is conveyed to the learner is unrelated to his or her performance, and is delivered by an *instructor*. Instruction is the demonstration of how to suture skin, whereas feedback is *tailored*, delivered by a *coach*, and informs the learner what is wrong with his or her technique, and what should be done to correct it, with respect paid to the learner’s individual strengths, weaknesses and position on the learning curve. Feedback can be intrinsic – what the learner senses themselves during or after completion of the task – or extrinsic – feedback that comes from an external source (46). Such extrinsic feedback can either be proctored (human feedback), or independent (for example, from a computer), such as the performance metrics presented by virtual reality (VR) and hybrid laparoscopy simulators (41).

When studied it is important to examine the effects of feedback on both skill acquisition and retention. In 2017, 36 medical students were randomised to practice a series of basic laparoscopic skills tasks on either a video-trainer or a hybrid simulator that provide such feedback. Although both groups demonstrated equal skill acquisition and retention (at six weeks), performance of a high-fidelity VR laparoscopic cholecystectomy was significantly quicker in the hybrid group at both post- and retention-tests (47).

There are many examples in the literature suggesting that such self-study might be ideal for the purposes of clinical skill acquisition and retention, and that the addition of an instructor or coach to the learning experience may actually be a hindrance. However, these studies remain difficult to interpret. For example, as detailed earlier, Udani *et al*'s study of UGRA demonstrated a striking increase in the simulation time required to reach proficiency in the feedback group compared to the self-study group, with no differences between them in the post and 3-month retention tests (40). Feedback in the intervention group was limited to scoring aspects of the task as ‘satisfactory’ or ‘unsatisfactory’ without provision of specific guidance on how individuals should improve. This may be related to the task’s reliance on working memory – with satisfactory task completion being dependent on the memorisation of a sequence of steps – rather than fine motor skill that distinguishes it from other tasks such as laparoscopic surgery.

In another study, 36 medical students were randomised to either self-study (with computer feedback) or proctored (computer and human feedback) groups for the acquisition of both colonoscopic and laparoscopic psychomotor skills on a VR simulator. The proctored group received both instruction and specific feedback related to each individual’s performance. No differences were found between groups in either the number of repetitions or total time required to reach proficiency. After adjusting for covariates, the proctored group actually took longer to reach proficiency than the self-study group (48). In a follow-up study testing colonoscopic skill retention at a median of 4.5 months, there were no differences in proficiency scores or error rates between both groups, though performance was significantly better than at baseline (41). Criticisms include the small number of participants available at retention testing.

Shippey *et al* provided further evidence of the futility of human proctoring in SBME in their randomised study comparing novices’ skill acquisition in subcuticular suturing. 58 medical students watched a video demonstration of the task, and practiced either on their own with a video, with instructor feedback, or with neither. Post and 1-week retention tests were scored by blinded assessors. Only the proctored group improved both global and task-specific performance scores at the post-test assessment, yet it was only the video self-study group who demonstrated significant improvements in both subscale scores between the pre- and one-week retention tests (49).

While Shippey *et al*'s study may demonstrate the disparate effects that instructor feedback may have on skill acquisition and retention, the proctored group had a ratio of five participants to one instructor – the reality is that participants may feel forced to stop and listen to less relevant feedback being offered to others in their small group rather than continuing to practice in the short 30-minute training session that was offered. It is plausible that such uninterrupted practice may have therefore confounded the results at retention testing, and a more favourable teacher to student ratio may have yielded different results. Nonetheless, there is sound rationale for providing video access to the self-study group, allowing a more effective intrinsic feedback mechanism through comparison of participants' own performance with experts at their own will, as master learners would (49).

Nousiainen *et al* had conducted a similar study of skill acquisition in suturing and knot-tying in 24 medical students. These novices were randomised to three groups – one group watched a demonstration video once and continued with self-study, the second group watched the video as often as they wanted, and the third group had unlimited video access in addition to an expert (with a ratio of four students to one expert). Although there was an improvement in performance in all groups, there were no differences between them at the post-test or one-month retention test (50). It was suggested that the simplicity of the task rendered the addition of an expert superfluous, highlighting the potential relevance of task complexity to human instruction and feedback.

With this in mind, Bjerrum *et al* more recently published their study of 99 medical students randomised to instructor and computer feedback, versus computer feedback alone, in the more complex task of laparoscopic salpingectomy on a VR simulator. This task requires both motor skill and cognitive procedural understanding – learning was accelerated with human instruction, although both groups retained their skill similarly at the 6-month retention test

(51). Since instruction was delivered from a template and was not individualised, it might be difficult to interpret or apply these findings to a real-world curriculum.

So while a meta-analysis in 2014 concluded that “feedback” in SBME does translate to a moderate improvement of skill outcomes (52), there was a moderate but statistically insignificant effect of instructor feedback over simulator-generated feedback. However, study heterogeneity was high, and so in reality the effectiveness of instruction/feedback may depend on factors such as its content, the instructor/coach to student ratio and task complexity. Indeed, this meta-analysis also concluded that the timing of feedback can also affect learning outcomes, with summary feedback (delivered infrequently after several practice repetitions) being superior to concurrent feedback (delivered continuously during each repetition), and that this too may depend on the complexity of the task (52). The *guidance hypothesis* suggests that concurrent feedback may lead to over-reliance such that performance may degrade once it has been withdrawn (53). *Cognitive load theory* also suggests that concurrent feedback can be overwhelming during a procedural skills session (54).

Although the evidence may not be strongly in favour of human proctoring overall, it is difficult to omit its inclusion in any training curriculum. Simulators and skills training curricula may not capture every parameter pertinent to a safe and competent performance. Poor habits using laparoscopic instruments may contribute to risk but may go unrecognised by a simulator (for example, introducing laparoscopic scissors into the abdomen with blades open, directed downwards towards bowel). Studies of instruction and feedback are not going to be sufficiently powered to detect any effect on real complications when these rarely occur. It could be argued, therefore, that human proctoring allows the deconstruction of a task during the *integration* phase of Fitts & Posner’s 3 phases of learning, rather than skipping to the *automation* of potentially poor habits and technique (31,44), with this conscious integration potentially prolonging the time taken to complete a task (44).

While most relevant simulator-derived metrics are time-based, Boyle *et al* suggested that human proctoring helps trainees perform a procedure *well* rather than *quickly* (55). Trainees may be taught to recognise an incomplete or poorly-performed subtask, to repeat it until it is adequate before continuing, rather than not recognising it at all and/or accepting it before simply moving on. In another example, Boyle *et al* described how incomplete dissection of the inferior mesenteric artery during a simulated laparoscopic colectomy may have resulted in better performance metrics in a group that was not specifically taught the clinical importance of proper dissection (46). There is also the suggestion that the presence of a human instructor or coach may counter against drop-outs from a training curriculum (56).

1.6.3 Distributed versus massed practice

A training schedule for learning can be defined according to the pattern of rehearsal. The *spacing interval* is the term often used to describe the time between these rehearsals. In a massed practice training schedule, the spacing interval is minimal or zero and all rehearsals occur within the same day. A distributed practice training schedule has a larger spacing interval of days, weeks or even months. These different training schedules are relevant to both the efficiency of learning and also the retention of new knowledge or skill following the final rehearsal. The interval between this final rehearsal and a test of retention is known as the *retention interval*.

It is now well known that distributed practice leads to greater retention in simple memory and verbal recall tasks (57,58). Spacing intervals result in a greater range of associated memories within each rehearsal, encoding a greater number of retrieval cues that can facilitate recall during retention tests, compared to massed practice learning. This theory of *encoding variability* is widely accepted to be responsible for the distributed practice effect in this context (57).

However, this rationale appears applicable only to *explicit* learning – where subjects are consciously aware of items that are to be recalled, and the manner in which tasks must be performed. Learning the psychomotor skills for performing laparoscopic surgery relies more on *implicit* learning – where subjects are not consciously aware of the sequences involved in the motor tasks they are learning. For example, identifying targets on a two-dimensional screen and activating specific muscle groups in order to effect the correct physical movement in the 3-dimensional (3D) environment. Indeed, functional imaging studies have identified distinct areas of the brain that are involved in implicit and explicit learning (59).

Still, there is accumulating data in favour of distributed practice in the context of learning these complex laparoscopic psychomotor skills. For example, 41 novices were randomised to either a massed practice group or one of two distributed practice groups to learn laparoscopic psychomotor skills on the MIST-VR virtual reality trainer. The group who practiced in four five-minute blocks with a spacing interval of only 2.5 minutes performed better during the latter stages of the learning phase and at the five-minute retention test, compared to the massed practice group who trained continuously for 20 minutes (60). In another study with a much greater spacing interval of one week, 38 junior surgical residents were randomised to either the massed or distributed practice group for training to perform a microvascular anastomosis. The distributed practice group outperformed the massed practice group both at the one-month retention test and on the transfer test where subjects performed an in-vivo microvascular anastomosis on anaesthetised rats (61). Similar results have been found in other studies comparing the effects of distributed and massed practice in surgical skills training (62–65).

The *memory consolidation hypothesis* is believed to be responsible for the effects of distributed practice on learning (66), and sleep may play an important role in this, allowing unstable memory forms to be stabilised (67). Furthermore, continued practice of motor tasks during a *consolidation* phase may interrupt this process and so may be detrimental to overall learning (66).

In an attempt to identify the boundaries and influencers of the distributed practice effect, Donovan *et al* conducted a detailed meta-analysis that separated studies into groups according to task type (mental versus physical), overall complexity (defined according to the number of choices, distinct behaviours and the degree of uncertainty within a task) and the spacing interval. It was found that for both acquisition and performance retention, the effect of distributed practice was more pronounced for tasks of lower overall complexity and high physical rather than cognitive requirement (68) – this is a description that may apply to the acquisition of psychomotor skills in laparoscopic surgery.

There are still only a handful of small studies that investigate the effect of distributed practice on the acquisition of surgical skill, and while they support it, little progress has been made in identifying the optimal spacing interval, which has also been shown to vary according to task type (68).

De Win *et al* attempted to identify the optimal spacing interval in 145 novice medical students being trained in laparoscopic basic skills and suturing by separating them into six groups of different spacing intervals, including two massed practice groups that mirror the practice regimen of traditional laparoscopy courses. Immediately after the training sessions were completed, skill acquisition was significantly better in those who trained once per day, compared to the massed or weekly training schedules. This advantage is still seen at 1 month, but the difference between daily and weekly practice is lost by 6 months, although the negative effect of massed training is still seen (69).

The benefits of daily distributed practice over weekly training, however, was not demonstrated in a study of 24 surgical interns that were randomised into these two groups to practice end-side vascular anastomosis. There was no difference in skill acquisition immediately after completion of training, or at the 4-month retention test (70). In 20 physicians randomised to either one-day or a weekly distributed practice schedule in bronchoscopy simulation, there was no difference in skill acquisition nor skill retention at four weeks (71).

Although it appears that distributed practice is superior to massed practice for both skill acquisition and retention, there is insufficient evidence to clearly identify the optimal spacing interval for learning laparoscopic psychomotor skills. Until more research has been completed to identify the influencers of the distributed practice effect, such training schedules are likely to be designed around cost and logistical limitations of maintaining the ideal open access training environment, with spacing intervals further impacted by the conflicting commitments of busy surgical trainees.

1.7 The Role of Simulation in the Acquisition of Laparoscopic Psychomotor Skill

Laparoscopic skills simulators provide the opportunity for trainees to develop their laparoscopic psychomotor skills in a safe environment, without the potential for patient harm. The suggested features of the ideal laparoscopy simulator can be divided into *operational* and *educational*, and are presented in Table 3.

Suggested Features of the Ideal Laparoscopy Simulator
Operational Features: > Cheap/affordable equipment > Reliable equipment, insusceptible to malfunction or breakage > No ongoing costs, such as servicing, maintenance and consumables > Intuitive/easy set-up > Portable, allowing easy translocation between training venues > Ethical
Educational Features: > Provides metrics for goal-oriented training > Possesses: - face validity (resembles reality) - construct validity (distinguishes between expert and novice performance) - predictive validity (practicing leads to translation of skills to the real environment) - Allows retrospective performance analysis

Table 3 – Suggested Features of The Ideal Laparoscopy Simulator.

In general, laparoscopic simulators can be divided into three types, each with their own advantages and disadvantages. *Video trainers* comprise a physical box together with a video camera and screen, allowing instruments to be passed and the trainee to practice with real accessories. Typically, these are relatively inexpensive and are easy to set up. They are often portable, allowing trainees to take them home for a period of time, as has been the case for some surgical trainees in the London Deanery (72).

VR simulators utilise computer-generated 3D images – rather than a camera and physical space to stage real psychomotor exercises using real objects, a computer generates a virtual display of a much wider range of practical scenarios. In doing so, the computer must instantly translate and perfectly coordinate the movements of the instruments and camera to the observed movements on the screen. This requires computer hardware and specialist software, often elevating the cost quite dramatically.

The major benefit of such a simulator is the provision of computer-generated objective performance metrics – measures of either time, movement or performance errors/accuracy. Not only does this allow a detailed retrospective analysis of performance, it also lends itself towards goal-directed learning. Furthermore, it could theoretically allow objective standards to be created that pertain to an individual's psychomotor skill. Creating these standards of technical skill is a formidable task in the context of surgery, and has not yet been accomplished, despite it being a necessary element in the quest to reduce inter-operator variability and guarantee high performance within a group of individuals (i.e. surgeons).

The third type of laparoscopic skills simulator is the hybrid simulator – a video trainer with specialist hardware attached that measures motion parameters. Whilst the cost is lower than that for VR simulation, certain parameters (for example, ideal path length and error scores) can be difficult to measure. Furthermore, resetting a task takes time, and setting up procedural tasks (such as laparoscopic cholecystectomy) brings identical challenges to those faced by a pure video trainer.

As regards VR simulation, there are various simulators available. These can be low- or high-fidelity, depending on the quality of the displayed images and the provision of haptic feedback

– a carefully calibrated force that is fed back to the operator, simulating the ‘feel’ of tissues in order to further enhance realism. The LAP Mentor (Simbionix, 3D Systems Corporation, CO, USA) is one such high-fidelity laparoscopy simulator that delivers haptic feedback. A wide range of skills modules can be applied to it that simulate psychomotor skills exercises and a wide range of full procedures, such as LA.

1.7.1 The Role of Haptic Feedback in VR Laparoscopy Simulation

Haptic feedback in laparoscopy is the input of physical sensations from the tissues to the hands of the operator via the instruments. Compared to open surgery, such feedback during laparoscopy is naturally attenuated and surgeons are less able to instinctively use their sensation of touch to identify and distinguish between structures and to regulate forces on the tissues. Video trainers may simulate laparoscopy but the haptic feedback is very real. However, in VR no haptic feedback exists unless it is simulated by way of technology (as in the LAP Mentor) and so it may be limited in its fidelity and, therefore, affect its validity. The advantage of video trainers in this regard is demonstrated in a study of 19 gynaecological residents randomised to either perform three laparoscopic tasks first on a video trainer and then the SIMENDO VR trainer (Box-VR), or the other way around (VR-Box). In the tasks where force plays little role, no differences in performance were demonstrated. However, in the task involving stretching an elastic band between two rings, the Box-VR group performed significantly better on the video trainer. Further, performing the elastic task on the video trainer first improved metrics subsequently in VR, but not the other way around (73). This suggests that haptic feedback may play an important role in training, and its absence in VR has the potential to affect the translation of skills into the real OR.

In fact, haptic feedback in VR laparoscopy may have the potential to either enhance or impede training, and as such may plausibly improve or even worsen performance in the OR. Irrelevant stimuli may complicate trainees’ learning (74), or if the fidelity of haptic feedback is poor then trainees may learn to operate with haptic feedback that is not actually experienced in the real environment. For example, experiencing excessive haptic feedback in VR may feasibly lead to a trainee exerting too much force on real tissues as the cues to relieve tension to which the trainee has become accustomed are absent. Indeed, the role of haptic feedback in the regulation of forces upon tissues has been highlighted in a study comparing a telerobotic task with and without haptic feedback – its absence saw the exertion of greater forces and more errors (75). Another demonstrated significantly greater maximum stretch damage on the 3D LapSim suturing task when haptic feedback was disabled (76). Still, does this mean that training in minimally invasive surgery without haptic feedback makes trainees prone to error in the real environment, or do they eventually adapt using visual cues that actually enhance real operating performance?

As alluded to, the published literature seems to suggest that the effect of haptic feedback on training depends on the task. There are those where haptic feedback is key to the task, such as tightening a knot, and others where they are not relevant, such as those requiring only eye-hand coordination. For example, 10 novices performed the peg transfer and cutting FLS exercises on a VR simulator with and without haptic feedback. Only the latter task, which required the exertion of tension on the material, demonstrated significant improvement in completion time when haptic feedback was enabled (77). On the other hand, 33 novices were randomised to either control, haptic feedback or non-haptic feedback groups. Training on the LAP Mentor BS tasks was followed by the laparoscopic cholecystectomy procedural module, and only minor differences in performance were found between haptic feedback and non-haptic feedback trained participants (78).

A systematic review was published in 2018 to examine the literature regarding the value of haptic feedback on laparoscopy training in VR. It was concluded that more complex tasks benefit from haptic feedback, where it shortens the learning curve and leads to more consistent performance among trainees, and novices tend to benefit the least (79). However, it was acknowledged that the quality of evidence was weak – there was significant heterogeneity in study design and sample sizes were very small. Furthermore, different simulators utilise different haptics technologies, and the same models may even receive updated haptics. This has contributed to an abundance of contradictory results, which makes it difficult to draw any firm conclusions.

The uncertainty of the effects of haptic feedback is compounded by the limited fidelity afforded by current simulation technology. For example, in one study surveying 22 laparoscopic surgeons, the LapSim's laparoscopic suturing task was rated as significantly more realistic with haptic feedback compared to without (which is hardly surprising, since any haptic feedback is likely to be better than no feedback at all), but ultimately between 4 and 13 participants rated each of three of the module's haptic feedback components as 'totally unrealistic' (76).

There are also particular aspects of real haptic feedback that is completely overlooked in current simulation technology. In real laparoscopic surgery, there is friction between the instruments and the ports, which is modulated according to the torque applied through the abdominal wall, and this will vary according to the angle the port was placed through it and the port's position relative to the operative field. This might be a crucial element of haptic feedback, since an operator is often unable to distinguish between the haptics from the actual tissue being operated on and these abdominal wall effects (80). Current simulation technology focuses only on the former.

The question remains – is it worth it? Haptic feedback confers significant additional costs and this may be better spent elsewhere. However, if skill acquisition can be accelerated, such as in more advanced trainees learning complex tasks that may include knot-tying, then it may be justifiable (74). It may also be worth considering the extent to which an increase in fidelity of a simulator as a whole (its face validity) might impact on a trainee's motivation to participate in a VR curriculum.

Future research in this area should concentrate on the effect of training with haptic feedback on the performance in the real operating environment. Further development may also see the inclusion of force-based metrics in VR curricula, as it has been suggested that force may be a better indicator of performance than classical time and position metrics (81). It has also been suggested that visual feedback of the forces being exerted in the VR environment may replace more expensive haptic feedback (79), though there remains the danger of training participants in how to use a simulator rather than how to improve real laparoscopic skills.

1.8 Validating Simulators

By definition, simulation is not indistinguishable from reality, and so it must not be taken for granted that the skill purportedly being taught from its use is the skill intended to be learned. It is broadly understood that the degree to which these align defines the validity of the instrument. Validation of any simulator (or related curriculum) is, therefore, essential, and its method of validation must be robust if it is to be trusted as a tool for learning.

The theoretical framework by which simulators and their tasks have been validated in the past have included terms that were originally proposed by the American Psychological Association (APA) in their standards of assessment published in 1954 (82). Initially, *content*, *concurrent*, *predictive* and *construct* validity featured in these recommendations, which evolved to also include *criterion-related* validity (**Error! Reference source not found.**) (83). The word 'validity' itself was originally used to describe 'the degree to which a test or examination measures what it purports to measure.' (84) However, more recently the term has begun to fall out of favour due to disagreements in how to apply it and the breadth of its meaning, making it difficult to fully understand (85).

Types of Validity	Definition
Construct validity	The extent to which a test measures what it purports to measure.
Content validity	The extent to which a test measures facets that are directly relevant to the construct of interest.
Criterion-related validity (2 sub-types):	The extent to which a test can predict an alternate measure.
Concurrent validity	The extent to which a test agrees with a different valid test.
Predictive validity	The extent to which the outcome of a test predicts future performance.

Table 4 – Types of validity, according to the American Psychological Association Theoretical Framework (83).

An alternate framework by Samuel Messick was proposed in 1989 (86) and adopted by the APA in 1999 (87), the American Educational Research Association (AERA) and the National Council of Measurement in Education (NCME). Abandoning all of the aforementioned validity terminology, the Messick Contemporary Framework categorises validity evidence according to five sources (88):

1. **Content:** evaluation by expert opinion, and may include questionnaires and Delphi methodology;
2. **Response process:** evaluation of bias in assessment, which may include efforts in standardisation, blinding of raters and assessments;
3. **Internal structure:** evaluation of the reliability of scoring, such as inter-rater reliability;
4. **Relations to other variables:** correlations between scores and other independent measures;
5. **Consequences:** potential and actual consequences of the assessment, such as its utility in defining a standard.

The existence of this newer framework might lead to criticism of studies that incorporate the older APA terminology. For example, the current authors argue that differences in the performance between experts and novices, which define construct validity, suggest the tasks and metrics that are most relevant for inclusion in any simulation curriculum. However, there is a risk of *construct-irrelevant variance*, where distinctions in performance are not related to a gap in relevant skill, but are due to irrelevant factors such as familiarity with the simulator itself, or any other unidentified aspect that bears no relevance to the skill ostensibly being tested or developed (89). It is for this reason that some have even argued that construct validity is only ever useful when absent, by identifying aspects that are almost certainly useless (90). Construct-irrelevant variance is an important consideration and is the reason why a study that wishes to validate a curriculum should exclude those who have had prior experience on the simulator upon which the curriculum is based. To go one step further by excluding those with any prior simulator experience may risk leaving very few eligible participants.

In a review of 51 eligible articles discussing validity and laparoscopic simulators (thereby including studies in respect to both training and assessment), Korndorffer *et al* argue that there is little use in simulators as a test to discriminate between experts and novices, using the divergence between a college math professor and a first-grade student as an example of irrelevance and inapplicability (89). It is also implied that using experts would be unrealistic for setting standards. However, the utility of a construct-valid metric would in fact depend on the context, specifically: are the discriminatory metrics used to set standards for training or for assessment? Are the groups that are found to be distinct in the measured construct relevant and applicable to the population intended to be assessed or trained? And finally, are expert level metrics truly unrealistic, or are they in fact desirable and realistic?

The distinction between intention to test or train has been notably absent from the discourse, but it is highly relevant. If a tool is to be used for testing, then the consequences of passing must correspond with the standard that is set, which is not usually a warrant for independent practice of real-world expert-level activities. Thus, using experts in a construct validity study designed to set standards for a *test* is likely to be inappropriate, as Korndorffer highlights. However, in a *training* environment, with a clearly-defined construct and a highly-specific learning domain (for example, laparoscopic psychomotor skill), setting expert-level standards may well be achievable. Indeed, for a curriculum where construct validity has been determined using a novice group that is exactly representative of the population for whom the curriculum is designed, the current authors believe that using ‘experts’ (actually more acceptably termed ‘experienced surgeons’) and ‘novices’ in its validation is entirely appropriate. The alternative on setting standards and curriculum validation might be to take Korndorffer’s advice, that ‘for a simulator to be used for training, little more than expert opinion…is needed’ (89).

Those who do not distinguish between the testing and training environment have levelled criticism against investigations that use an ‘outdated validity framework’ (88) – regardless of whether a study focuses on testing or on training, despite the Messick Contemporary Framework explicitly referring only to the former. Indeed, older frameworks are highly pervasive in the simulation literature. In Korndorffer *et al*’s aforementioned paper from 2010, every single reviewed article used older nomenclature regarding the types of validity and 23% had only partly used the Messick Contemporary Framework (89). In 2014, Cook *et al* similarly surveyed technology-enhanced simulation literature, and of 217 eligible studies, found that only 3% had referenced it (90). Later, in 2017, Borgersen *et al* found this figure was 6.6% of 498 studies (88), though there was a general trend of gradually increasing utilisation.

Suggested reasons for the infrequent reference to the Messick Contemporary Framework have included a lack of awareness (90) or that there is often a lack of personnel in research teams that have the necessary specialist educational background (88), suggesting that it may simply be too complicated compared to the pre-existing framework. *Consequences* and *response process* are examples of sources that are 'notoriously difficult to understand' (90). Others have defended the use of the older framework by pointing out that the Fundamentals of Laparoscopic Surgery (FLS), a committee set up by the Society of American Gastrointestinal Endoscopic Surgery (SAGES), developed a widely-recognised curriculum using the old framework, yet it is endorsed by the American College of Surgeons (ACS) (91,92).

The methodology used for validating tools for assessment and training using the older framework is at least partly transferable, and indeed Borgersen *et al*'s large survey found this to be the case in at least 80% of studies (88). Thus, rather than limit the use of terminology associated with older frameworks, it may be more constructive to understand where the evidence fits in the Messick Contemporary Framework, which may simply help to provide an alternate perspective on what it means for a simulation tool to be 'valid', and while considering the implications of whether the tool is used for training or assessment purposes.

1.9 Project Aim

Using the LAP Mentor and its LA module, this project aims to produce and implement an evidence-based VR curriculum for the acquisition of the psychomotor skills necessary to perform LA. The curriculum shall be aimed at trainees naïve to laparoscopy, who wish to pursue a career in surgery.

1.10 Study Objectives

In pursuit of the project aim, the primary objectives of this study are:

1. To test the hypothesis that the LAP Mentor LA module demonstrates construct validity;

2. To produce an evidence-based VR curriculum for LA using the LA module;

3. To implement the evidence-based VR curriculum with a cohort of novice trainees.

The secondary objectives are:

1. To assess the face validity of the LA module;

2. To demonstrate the learning curves of those naïve to laparoscopy in their practice of:
    a. the LAP Mentor psychomotor skills tasks pertinent to LA;
    b. the LA module's five individual procedural tasks (together these make up the full appendicectomy);
    c. the LA module's full unguided LA procedure;

3. To obtain qualitative feedback from participating trainees during the curriculum's implementation.

2 Methodology

2.1 Overview

The project was divided into three parts. Each employed a unique strategy to address the project's objectives and are as follows:

2.1.1 Part 1

Objective: Assessment of construct & face validity

Overview: Prospective, randomised study to compare expert and novice performance data from the Basic Skills (BS) and LA modules of the LAP Mentor. A questionnaire was employed to assess the face validity of the simulator and the LA modules.

2.1.2 Part 2

Objective: Construction of the VR LA curriculum

Overview: Construct validity results from Part 1 informed the construction of the VR LA curriculum.

2.1.3 Part 3

Objective: Implementation and evaluation of the curriculum

Overview: The new curriculum was implemented and participant diaries were analysed in order to study participants' views of the curriculum.

2.2 Study Protocol

The study was carried out at the Centre for Screen-Based Simulation (CSBS) at the Royal Free Hospital (London, United Kingdom). Recruitment began on 19th June 2013 and data was collected between 17th July 2013 and 30th March 2021 across three distinct parts of the study.

Part 1 involved recruitment of 30 junior trainees inexperienced in laparoscopy, and 27 experienced laparoscopic surgeons, with the sample size determined according to a power analysis. During induction, informed participant consent was obtained and each participant was randomised to one of three sub-groups of tasks to be performed laparoscopically and unassisted on the Lap Mentor VR simulator. Group A performed all 9 BS tasks, group B performed the 5 procedural tasks of the LA module, while group C performed the full LA procedure. The performance of novices' first two repetitions of each task was compared to those of the experienced group in order to test construct validity, and novices completed 8 further repetitions to establish the presence of a learning curve. These results informed the second part of the study.

During Part 2, the authors reached a consensus for the design of a new LA curriculum, founded upon a strong theoretical and evidence-based framework. This included agreement for the inclusion and proficiency criteria of specific tasks and metrics of the BS and LA modules.

Part 3 of the study was a prospective, qualitative evaluation of the curriculum as completed by trainees who were at the beginning of their surgical career. Each participant completed a participant diary, which underwent thematic analysis and formed the primary outcome data. Participant recruitment was terminated once thematic saturation was reached according to a validated method. Secondary outcomes were quantitative, and included data such as the number of repetitions and sessions required to complete the curriculum successfully.

All data collected during the study were anonymised and collated into an Excel spreadsheet (version 16.x, Microsoft, Redmond, WA, USA) and stored on a secure University College London computer, and analysed using SPSS (version 25, IBM, Chicago, IL, USA).

2.3 Setting

Data collection and curriculum implementation took place at the CSBS at the Royal Free Hospital (London, United Kingdom). This hospital is an 839-bed tertiary referral centre and is part of the Royal Free London NHS Foundation Trust. The Royal Free Hospital is also a campus for University College London Medical School (UCLMS)

The CSBS is located in the lower ground floor of the hospital, itself easily accessible by public transport, directly adjacent to the medical school campus infrastructure. It has a 24-hour seven-days-a-week open access policy and is directed by the Head of the Screen-Based Simulation Centre and Associate Professor of Surgical Science (Senior investigator). The CSBS consists of four dedicated simulation rooms: team-based simulation, endovascular, endoscopy and laparoscopy rooms. There is also one large debriefing room and one seminar room. The dedicated laparoscopy room of the CSBS houses three LAP Mentor simulators for both the delivery and research of VR curricula (93).

2.4 Equipment

The LAP Mentor is a high-fidelity VR laparoscopic simulator developed by Simbionix (3D Systems Corporation, CO, USA). It comprises a wheeled base and height-adjustable housing for the computer, a simulated camera, two laparoscopic instruments together with the hardware responsible for the delivery of haptic feedback to these instruments. Mounted above this is a flat-panel screen and keyboard. Portability is limited due to the size and weight of the machine.

The CSBS houses three of the second version of this simulator, the LAP Mentor II (**Error! Reference source not found.**). This simulator is installed on a housed desktop computer with an Intel® Core™ i7-3770 3.40 GHz Central Processing Unit and 8 GB RAM, which runs on the Windows 7 Professional 64-bit operating system (Microsoft, Redmond, WA, USA). From here, a series of training modules can be run.

Figure 2 - LAP Mentor II simulator. *Based in the laparoscopy room of the Centre for Screen-Based Simulation, Royal Free Hospital, London, United Kingdom.*

There are three basic laparoscopic skills modules, and are as follows:

1. Basic skills
    a. Camera manipulation (0°)
    b. Camera manipulation (30°)
    c. Eye-hand coordination
    d. Clip applying
    e. Clipping and grasping
    f. Two handed manoeuvres
    g. Cutting
    h. Electrocautery
    i. Translocation of objects
2. Laparoscopic essential tasks
    a. Peg transfer
    b. Pattern cutting
    c. Placement of ligating loop
3. Basic and advanced Suturing

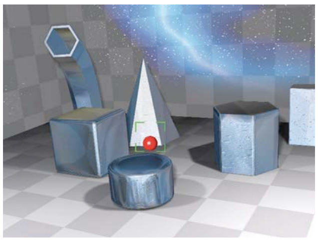
There are also 11 procedural laparoscopic training modules:

1. Laparoscopic cholecystectomy
2. Laparoscopic appendicectomy
3. Ventral hernia
4. Inguinal hernia
5. Gastric bypass
6. Sigmoid colectomy
7. Nephrectomy
8. Essential GYN procedures
    a. Tubal sterilization
    b. Ectopic pregnancy – salpingostomy
    c. Ectopic pregnancy – salpingectomy
    d. Prophylactic oophorectomy

9. Hysterectomy
10. Vaginal cuff closure
11. Thoracic lobectomy

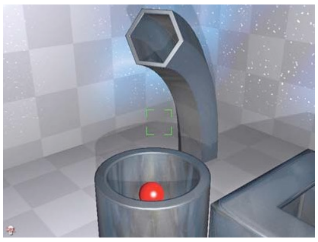
The BS and LA modules were studied and formed the new VR LA curriculum. The tasks that comprise the BS module are outlined in Table 5. The LA module consists of five procedural tasks (Figure 3) and the full procedure (FP). Each procedural task concentrates on a specific aspect of the full procedure, with text and coloured indicators to guide the operator. The FP does not include any such guidance. In total, the basic skills and appendicectomy modules report on 125 and 58 performance metrics, respectively. *Number of movements* and *path length* are both provided for each right and left instrument, but no *total* is provided, and so this was calculated and included in all subsequent analysis. Therefore, the BS and LA modules each yielded a total number of 143 and 70 performance metrics, respectively.

1. Camera navigation (0º)



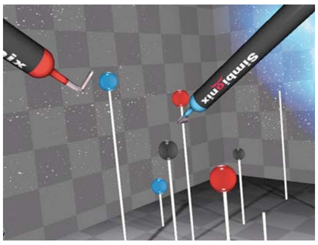
A 0 degree camera is used to take photographs of a red ball in varying positions.

2. Camera navigation (30º)



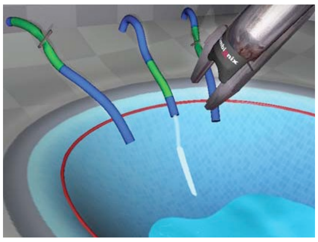
A 30 degree camera is used to take photographs of a red ball in varying positions.

3. Eye-hand coordination



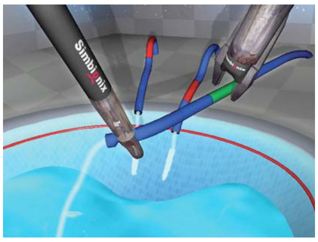
Each coloured ball must be touched with the instrument of the same colour.

4. Clipping



A clip must be applied to each pipe within the green-marked zone.

5. Grasping & clipping



Each pipe must be grasped with tension applied in order for the red zone to change colour to green. Then a clip must be applied to this green zone with the clip applier in the alternate hand. This must be completed with 9 pipes before the pool fills with water that reaches the red line.

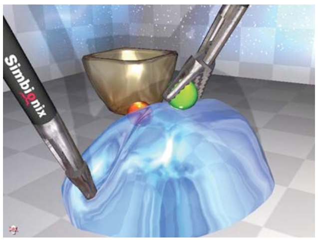
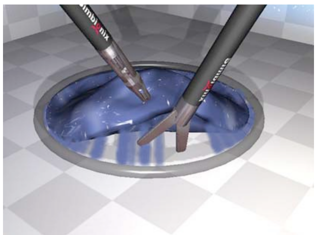
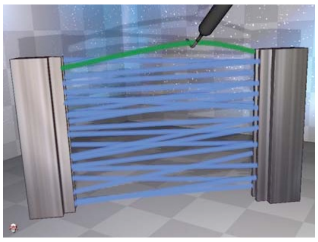
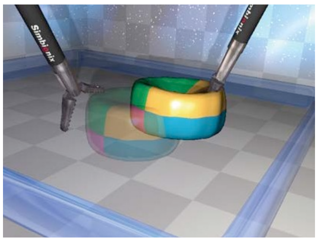
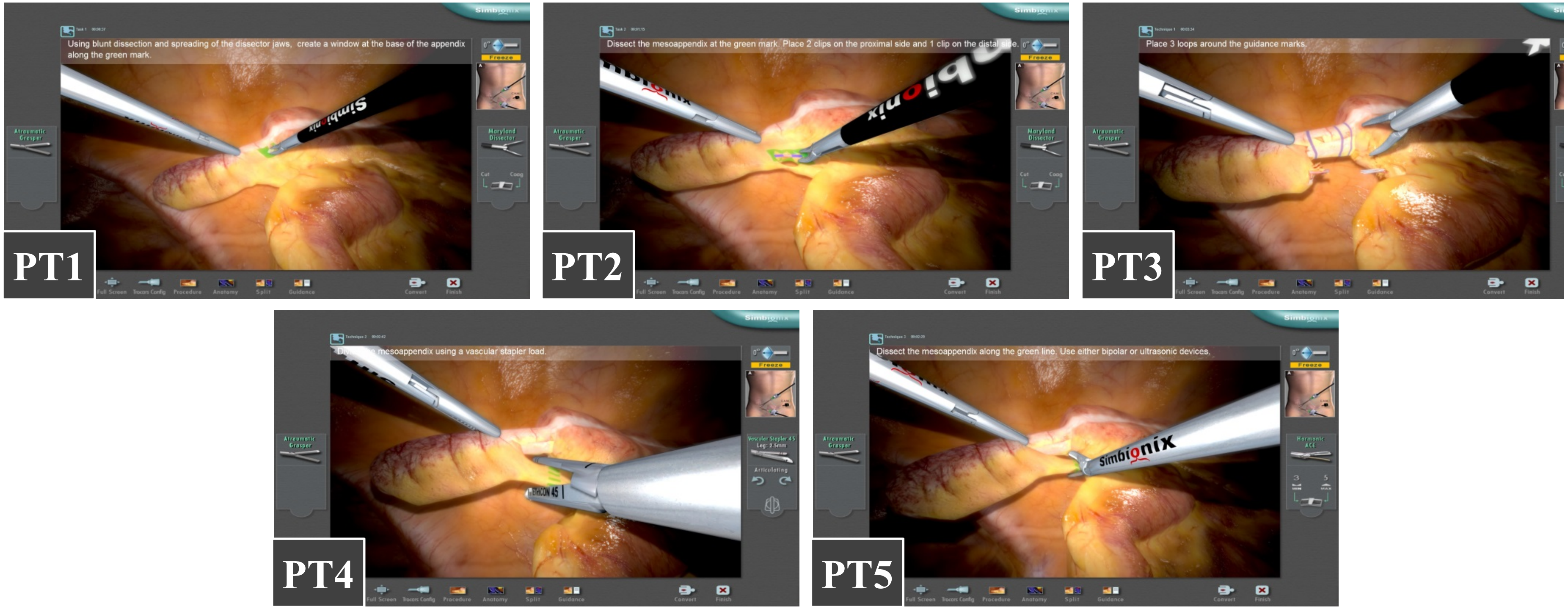
6. Two-handed manoeuvres		One hand must retract the jelly-like substance in order to retrieve a virtual marble, which must then be placed in the container. This must be repeated until all marbles have been retrieved.
7. Cutting		One hand must create tension between a circular structure and its peripheral attachments, while the alternate hand uses the scissors to cut them, so as to free the structure.
8. Diathermy		The hook diathermy must be used to divide each band, using the pedal to apply energy. Bands turn green, indicating the order in which to divide them, while avoiding inadvertent injury to blue bands.
9. Object translocation		Both hands are used with graspers to orientate and place several multi-coloured objects in the virtual shadow with the indicated orientation.

Table 5 - Screen Capture Images of the Laparoscopic Basic Skills Tasks. *Images reproduced from Sinitsky* et al *(2012) (94).*



Task 1 00:00:37
Using blunt dissection and spreading of the dissector jaws, create a window at the base of the appendix along the green mark.

Task 2 00:01:15
Dissect the mesoappendix at the green mark. Place 2 clips on the proximal side and 1 clip on the distal side.

Technique 1 00:03:34
Place 3 loops around the guidance marks.

Technique 2 00:02:42
Di mesoappendix using a vascular stapler load.

Technique 3 00:02:29
Dissect the mesoappendix along the green line. Use either bipolar or ultrasonic devices.

Figure 3 - Screen Capture Images of the Laparoscopic Appendicectomy Procedural Tasks. *Colour codes guide trainees throughout the surgical steps. PT1: Dissecting the mesenteric window; PT2: Dissecting the mesoappendix and clipping the artery; PT3: Clipping the artery and ligating the appendix using a ligating loop; PT4: Division of the mesoappendix and base of the appendix using a stapler; PT5: Control of the artery using energy. PT = Procedural Task.*

2.5 Part 1 – Assessment of Construct & Face Validity

Construct validity informed the inclusion of tasks and individual metrics in Part 2 of the project (construction of the VR curriculum). To achieve this, a prospective, randomised trial was conducted, comparing novice with expert performance for each individual metric across all LAP Mentor BS tasks, and across the five procedural and FP tasks of the LA module. Participants who were randomised to perform the LA procedural or FP tasks then provided an assessment of face validity.

2.5.1 Sample

In this experiment there was a ‘novice’ cohort of participants (naïve to laparoscopy) and an ‘experienced’ cohort of experienced laparoscopists. Definitions were based on a consensus from previously published research (95–98), and participants were recruited according to the criteria defined in Table 6. Although prior experience on the LAP Mentor precluded participation, experience on other laparoscopic simulators was permitted since this may otherwise have led to difficulties in recruitment.

The easiest method of designing a construct validity assessment would have been to compare the performance of a single novice group across all tasks on both the BS and LA modules, with that of experienced laparoscopic surgeons in a single group. However, a *warm-up effect* may exist, with practice early in a session likely to affect the performance of latter tasks. In order to minimise this possibility, while at the same time avoiding overly burdensome and unrealistic targets for participant recruitment, each group (novice and experienced) was divided into three subgroups. Subgroups N1, N2 and N3 consisted of novices performing exclusively either the BS, procedural, or FP tasks, respectively. Subgroups E1, E2 and E3 consisted of experienced laparoscopic surgeons, as defined according to the criteria in Table 6, and they were also allocated to perform either BS, procedural, or FP tasks, respectively.

Part 1 Selection Criteria

Inclusion criteria (novice):

Foundation trainee doctors, years 1 and 2;

Must have been primary operator for <10 laparoscopic procedures (excluding diagnostic laparoscopy);

Must have had primary assistant experience for ≥5 laparoscopic appendicectomies.

Inclusion criteria (experienced):

Consultant general surgeons, Specialty Registrars (ST3+);*

Must have been primary operator for >100 laparoscopic procedures (excluding diagnostic laparoscopy).

Exclusion criteria (all):

Prior experience on the LAP Mentor simulator (experience on other laparoscopic simulators was permitted).

Table 6 – Part 1 Selection Criteria for Novice and Experienced Participants. **Specialty Registrars are at least postgraduate year 5, and ST3+ refers to the grade of the Higher Surgical Training programme for General Surgery in the United Kingdom (ST3 to ST8, where ST3 is year 1 of the programme).*

2.5.1.1 Power analysis/sample size calculation

Sample size was calculated according to the study's primary objectives, based on a two-tail test, where $\alpha = 0.05$ and power $(1 - \beta) = 0.80$, with an expected reduction in task completion time in the experienced group compared to the novice group of 30% based on previous studies (96,99,100).

This yielded a value of 8 participants per subgroup. Allowing for dropouts and anticipated challenges in recruitment, the sample size was set at a minimum of 10 for each novice subgroup and 9 for each experienced subgroup, totalling at least 30 novice and 27 experienced participants.

2.5.2 Recruitment & Randomisation Overview

In order to recruit novice participants, an email was sent to Foundation Year (FY) 1 and FY2 doctors across a range of hospitals within University College London Partners (UCLP), advertising the opportunity to participate in the study. Experienced participant recruitment was anticipated to be more challenging due to busier schedules and arguably a lower expected perceived return on the investment of their time (whereas novices may relish the opportunity to practice laparoscopy, no such novelty exists for experienced laparoscopists). As such, experienced laparoscopic surgeons were recruited by personal communication with consultants and trainees on the Higher Surgical Training programme for General Surgery across UCLP.

A double-blinded closed-envelope technique of participant randomisation was utilised during a standardised induction meeting for each recruited participant.

2.5.3 Participant Induction

Prior to commencement of simulator practice and data collection, each novice and experienced participant attended a mandatory induction meeting. This was standardised in order to avoid the introduction of bias and maintain methodological transparency, and so each induction meeting followed an identical checklist (appendix i), which formed part of an *induction pack* that was placed inside each of 60 pre-prepared unmarked envelopes. These were separated into two groups (novice – subgroups N1 to N3, and experienced – subgroups E1 to E3). At the start of each meeting, all remaining envelopes in the relevant group were shuffled and then one was selected. This informed the participant of their subgroup. The second part of the induction meeting consisted of a 15-minute briefing to cover the following areas:

1. Introduction to the study (appendix ii);
2. Pre-study questionnaire (appendix iii);
3. Signing of participant agreement and consent form (appendix iv);
4. Logistical issues;

5. Participant safety.

Following satisfactory completion of the briefing, a 20-minute simulator familiarisation session followed, which covered the following areas:

1. Introduction to the simulator;
2. Demonstration of the simulator and the relevant module, using standardised techniques agreed by the faculty (participants were able to practice for 10 minutes together with the instructor);
3. Opportunity to answer questions.

During the induction meeting, demonstrators were forbidden from offering any advice as to how to improve performance on the simulator, and it was not permitted to share any information as to which metrics may or may not be relevant.

2.5.4 Training Protocol

For all experienced and novice participants, each training session comprised one complete performance of all the tasks required of their group in consecutive order. The manner in which tasks were to be completed was standardised. For example, during the induction meetings it was stipulated that participants in the FP subgroups (N3 and E3) perform the procedure by dissecting the mesoappendix, double-clipping the appendicular artery before dividing it, and dividing the appendix base following the application of three ligating loops.

Novices were required to complete 10 training sessions, to commence on a separate day to their induction meeting. This is more than the number of repetitions by which novices have been able to reach expert-level proficiency in previous studies of VR laparoscopy modules (96,101–103). Experienced participants were required to complete only two training sessions, commencing on the same day as their induction meeting.

For the comparative assessment between experienced and novice performance, all participants performed their first two sessions on the first day of training. For the entirety of the study period, no more than two sessions could be performed on the same day, and no two sessions within one hour of each other. This was to ensure consistency regarding mental fatigue and consolidation of learning. After each performance, all participants were able to view their metric data.

In order to guarantee consistency, all participants were individually supervised by the same faculty member. This was to ensure adherence to the study protocol and to provide technical support with the simulator. Instruction and feedback were prohibited. Upon completion of all required repetitions, novices were provided with a certificate of completion.

2.5.5 Data Collection & Analysis

In order to demonstrate homogeneity between novice groups, the following demographic data was collected in the pre-study questionnaire that was completed during each induction meeting:

- age;
- gender;
- extent of previous laparoscopy experience;
- extent of previous simulator experience;
- hand dominance;
- experience with computer games;
- experience with musical instruments.

Performance data from the simulators were anonymized and collated into an Excel spreadsheet (version 16.x, Microsoft, Redmond, WA, USA) and stored on a secure University College London computer. Data was then transferred to SPSS (version 25, IBM, Chicago, IL, USA) for statistical analysis.

Data distribution for each metric across all simulator tasks was assessed using the Shapiro-Wilk test. For metrics where data was normally distributed, construct validity was assessed by comparing the mean of the first two repetitions in the novice group with that of the corresponding expert group using a two-tailed T test. For non-parametric data, the two-tailed Mann-Whitney U test was used to compare medians. The presence of a learning curve amongst novices was assessed for each construct-valid metric using the repeated measures ANOVA test for parametric data, and the Friedman test for non-parametric data.

Results were considered significant where $p < 0.05$. A metric was deemed construct-valid if performances were significantly different between experts and novices. A task was deemed construct-valid if it contained one or more construct-valid metric.

Upon completion of all required repetitions, all participants that were allocated to subgroups of the procedural and FP appendicectomy tasks (N2, N3, E2 and E3) completed an Opinio online survey (ObjectPlanet Inc, Oslo, Norway), to assess their perceptions of the realism and usefulness of the LA module. Answers were given on a 7-point Likert scale, one being the least useful/realistic/agree and seven being most useful/realistic/agree. Answers of four were deemed 'neutral', four to five as 'mild', five to six as 'moderate' and six to seven as 'strong'. Symmetrical definitions were used for answers between one and four.

A flow chart for Part 1 is outlined in Figure 4.

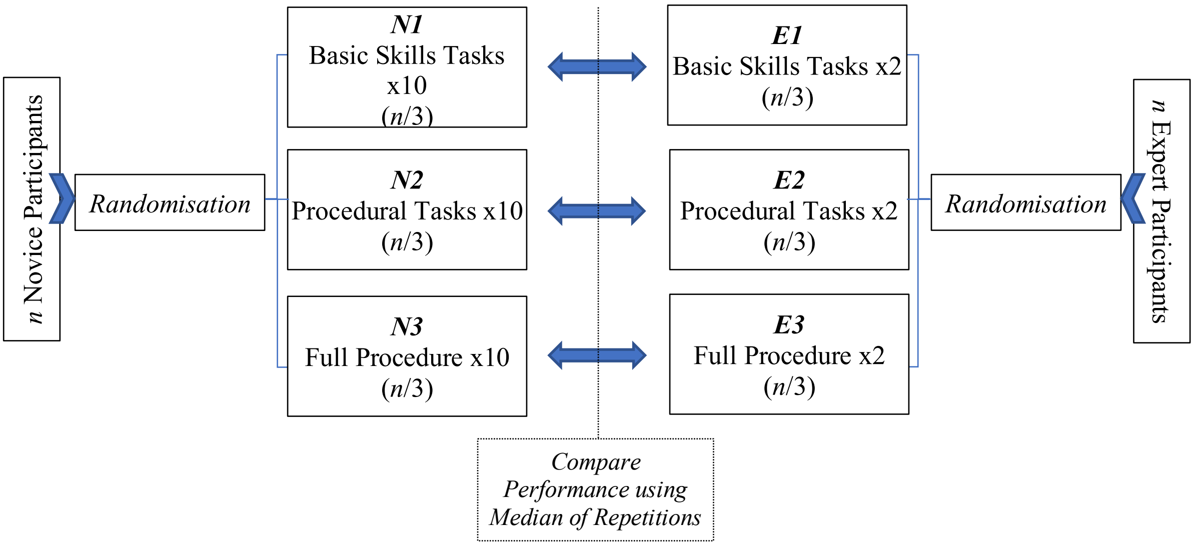


Figure 4 - Flow chart of the methodology for Part 1

2.6 Part 2 – Construction of the VR Laparoscopic Appendicectomy Curriculum

The VR LA Curriculum was designed using the methodology adopted by Aggarwal *et al* (96). The curriculum also heeded recommendations previously published elsewhere by the current research group (94).

The current research group initially discussed and agreed criteria for the design of the curriculum, which are presented in Table 7.

Curriculum Components	Design/Inclusion criteria
Metrics	> Must be construct-valid (excluding error scores, which are considered separately and included if clinically relevant) > Expert levels of performance must be attainable (i.e. demonstration of learning towards expert level proficiency) > Must be easy to comprehend, and able to inform the attempted improvement in subsequent performances* > Must be deemed clinically relevant*
Task	> Must be deemed clinically relevant* > Possesses one or more metrics meeting criteria set in the metrics component
Session format	> Distributed practice training schedule
Proficiency	> Goals directed based on the median of expert performance scores

Table 7 - Criteria adopted for the design of the curriculum. *A version of this table was previously published by the current author* (104). **Determined by internal consensus.*

Metrics and tasks were selected according to these criteria, and then a further discussion took place that finalised the design of the curriculum, regarding:

- the metrics to include/exclude;
- the experienced group's performance centile that would be set as the proficiency level to be passed, in order for participants of the curriculum to progress;
- how instruction and feedback would be delivered (content and timing);
- the duration and timing of practice sessions;
- the general design of a curriculum that would facilitate deliberate practice.

The available evidence for curriculum development and psychomotor skill acquisition was used to form the basis of this discussion, through which a consensus was reached. A final draft of the curriculum was then approved by all group members.

2.7 Part 3 – Implementation & Evaluation of the Curriculum

The objectives of Part 3 were to implement the VR LA curriculum among novice trainees, and to evaluate it by obtaining relevant feedback from participants regarding their experiences of completing it. As such, Part 3 was primarily qualitative in nature. Evaluation of the curriculum also included determining whether or not the proficiency criteria were realistic and achievable, and assessing whether or not trainees would voluntarily attend the CSBS regularly in order to complete a curriculum based on distributed practice. Consensus on these objectives and the methodology employed to meet them was obtained during two focus group sessions with experienced surgeons, research academics and educators.

2.7.1 Sample & Recruitment

The cohort that was identified for participation consisted of those individuals for whom laparoscopic skill development was relevant, who are at the start of their learning curve, and for whom opportunities for performing laparoscopic surgery on real patients were imminent. FY1 and FY2 doctors meet these criteria, and so emails inviting this cohort to participate were sent to the hospitals of UCLP.

In order to reinforce the above criteria and to ensure sample homogeneity, participation was limited to those who had been primary operator for less than 10 laparoscopic procedures (excluding diagnostic laparoscopy). Individuals with prior practical experience with the LAP Mentor were excluded.

2.7.2 Participant Induction

Applicants eligible to participate had a one-to-one induction meeting with the senior investigator at the CSBS that followed a similar format to that previously described for Part 1:

1. Introduction to the study;

2. Pre-study questionnaire (identical to that of Part 1);
3. Signing of participant agreement and consent form;
4. Logistical issues;
5. Participant safety.

During participant induction, the pre-study questionnaire implemented during Part 1 of the study was completed by each participant. Following satisfactory completion of the introductory briefing, participants were introduced to the simulator tasks during a familiarisation session. They underwent one-to-one teaching by the principal investigator on laparoscopy technique, as an introduction to the first part of the curriculum (psychomotor skills tasks).

2.7.3 Study Protocol

Following the induction meeting, each participant began the curriculum on a separate visit to the simulation centre. Participants were advised that each session must last no longer than 45 minutes, with no more than one session permitted in a single day.

Although the curriculum had initially instructed all tasks to be performed in a fixed order, all participants were granted permission to repeat tasks at their discretion. However, they were limited to the stage of the curriculum that they were in, and the requirement to meet the progression criteria in all tasks over two consecutive rounds had not changed. Apart from feedback sessions, all practice sessions were unsupervised.

Feedback sessions were provided by one of four faculty who are practicing laparoscopic surgeons who meet the criteria for experienced participants as set out in Part 1. These are one-to-one sessions that allowed the instructor to observe the performance of each participant and provide tailored feedback and demonstrations, while answering any questions. These sessions were mandatory following every fifth session if the participant had failed to progress to the next stage of the curriculum.

Feedback sessions also occurred following successful completion of each of the first two stages of the curriculum (BS and LA procedural tasks). As part of these sessions, each participant was introduced to the next stage of the curriculum by way of demonstration and hands-on familiarisation. Performance metrics during feedback sessions were excluded from analysis.

Upon completion of the final stage of the curriculum, the instructor then assessed the participant's performance using a modified version of the Global Operative Assessment of Laparoscopic Skills (GOALS) tool. The GOALS tool is a validated and reliable method for evaluating technical skill by experienced raters (105). However, performing laparoscopy with unsafe habits is not explicitly part of the scoring system, yet it may be useful given the drawbacks of a VR simulator in providing metrics that do not take these into account (as outlined earlier). As such, it has been added to the tool (Figure 5). The instructor provided written feedback at the end of this assessment and were advised that they may declare either that the curriculum had been successfully completed, or they could recommend that the participant returns to any stage of the curriculum for focused deliberate practice before repeating the final assessment. Participants were provided with a certificate upon successful completion of the curriculum.

	Depth perception
	1. Constantly overshoots target, wide swings, slow to correct
	2.
	3. Some overshooting or missing of target, but quick to correct
	4.
	5. Accurately directs instruments in the correct plane to target
	Bimanual dexterity
	1. Uses only one hand, ignores nondominant hand, poor coordination between hands
	2.
	3. Uses both hands, but does not optimize interaction between hands
	4.
	5. Expertly uses both hands in a complimentary manner to provide optimal exposure
	Efficiency
	1. Uncertain, inefficient efforts; many tentative movements; constantly changing focus or persisting without progress
	2.
	3. Slow, but planned movements are reasonably organised
	4.
	5. Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach
	Tissue handling
	1. Rough movements, tears tissue, injures adjacent structures, poor grasper control, grasper frequently slips
	2.
	3. Handles tissues reasonably well, minor trauma to adjacent tissue (ie, occasional unnecessary bleeding or slipping of the grasper)
	4.
	5. Handles tissues well, applies appropriate traction, negligible injury to adjacent structures
	Autonomy
	1. Unable to complete entire task, even with verbal guidance
	2.
	3. Able to complete task safely with moderate guidance
	4.
	5. Able to complete task independently without prompting
	Habits*
	1. Frequently awkward use of instruments, dangerous entry of instruments into abdomen, frequent off camera movements, awkward posture without use of roticulators
	2.
	3. Few awkward movements, uses roticulators but sub-optimally, appearance of moderate physical comfort, instrument travel into abdomen
	4.
	5. Comfortable posture, safe instrument travel into abdomen, natural use of roticulators, no awkward movements

Figure 5 – Modified Global Operative Assessment of Laparoscopic Skills (GOALS) assessment tool. **Denotes an additional domain that has not previously been validated and exists only for the purpose of this curriculum. Adapted from Vasilliou* et al (105).

2.7.4 Participant Diaries

During the induction meeting, each participant was provided with a participant diary. Participants were encouraged to comment on any positive or negative experiences regarding the simulator, curriculum and simulation centre. Participant diaries were anonymised prior to thematic analysis.

Since real-world implementation of the curriculum was intended to extend indefinitely beyond the end of the study period (subject to feasibility), and since Part 3 is mainly qualitative in nature, the sample size was dependent upon the number of participant diaries required to reach data saturation. This was measured according to the method published by Guest *et al* (2020) in reporting thematic saturation in qualitative research (106), with a desired New Information Threshold of less than 5%. At this point, data collection for Part 3 terminated and all participants who had completed the curriculum were included in the analysis.

3 Results

3.1 Part 1 – Assessment of Construct & Face Validity of the Basics Skills and LA Modules

3.1.1 Participants & Data

36 novices were randomized to groups N1 to N3 although one participant dropped out between induction and commencement of data collection, leaving groups N1 and N2 each with 12 participants and N3 with 11. One further participant in group N2 dropped out after the 6th repetition (data from repetitions one to six was included in the analysis), while the remaining 34 participants completed all 10 repetitions of their allocated task(s). Reasons for dropping out were not provided. The baseline characteristics of groups N1, N2 and N3 are presented in Table 8.

27 experienced laparoscopic surgeons were also recruited, with two dropping out between study induction and commencement of the first session (one citing a busy schedule, the other not providing a reason), leaving groups E1 and E2 each with eight participants, and group E3 with nine. All of the 25 remaining experienced participants attempted two repetitions of their allocated task(s).

	N1	**N2**	**N3**	**Total**
N	12	12	11	35
Male (female)	8 (4)	6 (6)	6 (5)	20 (15)
Participants with previous experience on VR simulator* (mean, mins)	2 (31.5)	1 (20)	2 (35)	5 (31)
Participants with previous experience on video ('box') trainer	4	5	6	15
Previous full laparoscopic procedures as primary operator	0	0	0	0
Previous experience as camera navigator	11	10	9	30
RH dominance (LH)	12 (0)	10 (2)	9 (2)	31 (4)
Plays musical instrument (hours per week)**	2 (1 – 1.5)	2 (1 – 3)	3 (0.5 – 1)	7 (0.5 – 3)
Plays computer games currently	7	4	7	16
Hours per week: 0	*5*	*8*	*6*	*19*
0-2	*5*	*1*	*6*	*12*
2-4	*1*	*0*	*0*	*1*
4-7	*0*	*2*	*0*	*2*
>7	*1*	*1*	*0*	*2*
Played computer games only in the past	4	7	4	15
Rarely played	*1*	*2*	*1*	*4*
Used to occasionally	*1*	*3*	*0*	*4*
Used to regularly	*2*	*2*	*3*	*7*
Never played computer games	1	1	2	4

Table 8 – Baseline characteristics of novice participants. **No participant will have had prior experience on the LAP Mentor, since this would have excluded any participant from the study – see table 4. **Number of hours per week instrument games played expressed as a range for only those that play at all. N1 = Psychomotor skills module; N2 = laparoscopic appendicectomy procedural tasks; N3 = full appendicectomy procedure; VR = Virtual Reality; RH = Right Hand; LH = Left Hand.*

Unexplained software failures led to missing data affecting a number of movement metrics in procedural task 4 (PT4) and PT6 (Table 9). Missing data from either or both of a participant's first two repetitions led to the exclusion of several participants from the construct validity analysis of these metrics (one to four novices and zero to six experts, depending on the metric).

Affected Metrics	Missing	Excluded
Metrics from Subgroup E1		
Task 4: Economy of movement - left instrument (%)	4	2
Task 4: Ideal path length of left instrument (cm)	5	3
Task 4: Ideal path length of right instrument (cm)	1	1
Task 4: Relevant path length - left instrument (cm)	5	3
Task 6: Economy of movement - left instrument (%)	5	4
Task 6: Economy of movement - right instrument (%)	1	1
Task 6: Ideal path length of left instrument (cm)	5	4
Task 6: Ideal path length of right instrument (cm)	1	1
Task 6: Relevant path length - left instrument (cm)	5	4
Task 6: Relevant path length - right instrument (cm)	1	1
Metrics from Subgroup N1		
Task 4: Economy of movement - left instrument (%)	19	1
Task 4: Ideal path length of left instrument (cm)	19	1
Task 4: Number of movements of right instrument	1	1
Task 4: Relevant path length - left instrument (cm)	18	0
Task 6: Economy of movement - left instrument (%)	43	6
Task 6: Economy of movement - right instrument (%)	3	1
Task 6: Ideal path length of left instrument (cm)	35	5
Task 6: Ideal path length of right instrument (cm)	11	2
Task 6: Relevant path length - left instrument (cm)	43	6
Task 6: Relevant path length - right instrument (cm)	3	1

Table 9 – Summary of missing data from Part 1. *Technical failure of unknown cause resulted in missing data that affected basic skills tasks 4 and 6 in both N1 and E1 groups. Missing data that affected either of a participant's first two repetitions led to exclusion of that participant's data for the purpose of construct validity analysis.*

3.1.2 Construct Validity & Learning Curves

Tasks 3, 5, 6, 8 and 9 of the BS module demonstrated construct validity. Eleven of 17 (65%) native metrics demonstrated construct validity for BS task 5, with zero to four native metrics demonstrating construct validity for all other BS tasks. Participants demonstrated significant learning in 17 of these 20 (85%) construct-valid metrics.

All five appendicectomy procedural tasks were construct-valid, with 33 of 48 (69%) of all native metrics in this group demonstrating construct validity and all but four of these demonstrated significant learning. In the FP, seven of 10 (70%) metrics demonstrated construct validity, with all of these construct-valid metrics demonstrating significant learning. Injury to the appendicular artery demonstrated construct validity only in PT3, although significant learning was not demonstrated.

No meaningful relationships related to dexterity were observed, owing to the relatively small sample size. Only four of 31 participants were left-handed and the remainder right-handed. All group A participants were right-handed.

Construct validity results for BS, procedural task and FP groups are presented in Table 10, Table 11 and Table 12, respectively. These include the non-native combined left and right-hand movement and path length metrics. Learning curves for selected tasks (those later included in the curriculum) are presented in Figure 6, Figure 7 and Figure 8, respectively.

Task	Metric	Description	Novices Median (range)	Experts Median (range)	ND? *P* value	CV? *P* value	LC? *P* value
1	Camera Manipulation 0 Deg	Accuracy rate - target hits (%)	95.5 (90.91 - 100)	97.7 (88.46 - 100)	0.000	ns	
		Average speed of camera movement (cm/sec)	8.8 7.65 - 12.09)	8.5 (7.82 - 10.07)	0.000	ns	
		Maintaining the horizontal view while using the 0° camera (%)	86.8 (71.59 - 98.9)	90.7 (54.52 - 99.28)	0.000	ns	
		Number of correct hits	10 (10 - 10)	10 (10 - 10)	0.000	ns	
		The time the horizontal view is maintained (±15°) while using the 0° camera	61.4 (50.67 - 80.68)	65.9 (53.49 - 78.28)	0.616	ns	
		Total Number of camera shots	10.5 (10 - 11)	10.3 (10 - 11.5)	0.000	ns	
		Total path length of camera (cm)	271.2 (233.82 - 369.22)	248.4 (214.73 - 342.99)	0.095	ns	
		Total time	82.8 (62.6 - 94.85)	80.6 (62.07 - 102.29)	0.004	ns	
2	Camera Manipulation 30 Deg	Accuracy rate - target hits (%)	95.5 (90.45 - 100)	93.2 (82.95 - 100)	0.000	ns	
		Average speed of camera movement (cm/sec)	8.4 (7.31 - 9.99)	8 (6.9 - 8.62)	0.000	ns	
		Number of correct hits	10 (9.5 - 10)	10 (9.5 - 10)	0.000	ns	
		Total Number of camera shots	10.5 (9.5 - 11)	10.8 (10 - 11.5)	0.000	ns	
		Total path length of camera (cm)	324.9 (281.32 - 508.03)	303.8 (217.45 - 417.99)	0.000	ns	
		Total time	87.8 (71.92 - 163.17)	102.1 (67.01 - 132.54)	0.000	ns	
3	Eye-Hand Coordination	Accuracy rate - touched targets (%)	100 (87.12 - 100)	100 (90.45 - 100)	0.000	ns	
		Average speed of left instrument movement (cm/sec)	3.1 (2.48 - 3.77)	2.8 (2.48 - 3.75)	0.054	ns	
		Average speed of right instrument movement (cm/sec)	3.5 (2.32 - 5.01)	3.1 (2.27 - 4.52)	0.144	ns	
		Economy of movement - left instrument (%)	60.7 (51.59 - 73.03)	72.1 (47.89 - 79.42)	0.813	ns	
		Economy of movement - right instrument (%)	48.1 (31.88 - 61.88)	64.4 (42.79 - 71.09)	0.694	0.003	0.000
		Ideal path length of left instrument (cm)	34.4 (27.51 - 36.73)	35.4 (31.89 - 39.7)	0.517	ns	

		Ideal path length of right instrument (cm)	32.1 (22.7 - 38.55)	31.5 (24.28 - 38.12)	0.064	ns	
		Number of correct hits	10 (9 - 10)	10 (9.5 - 10)	0.000	ns	
		Number of movements of left instrument	26.5 (17.5 - 28.5)	22 (15 - 28)	0.001	ns	
		Number of movements of right instrument	26.8 (19 - 37)	22.3 (15.5 - 42.5)	0.000	ns	
		Total number of movements	53.5 (37.5 - 65.5)	46.5 (31.5 - 68)	0.000	ns	
		Relevant path length - left instrument (cm)	53.5 (39.77 - 70.42)	49.8 (42.16 - 74.43)	0.331	ns	
		Relevant path length - right instrument (cm)	71.4 (36.6 - 117.02)	51.3 (40.86 - 71.77)	0.002	0.021	0.212
		Total Number of touched balls	10 (9 - 11.5)	10 (9.5 - 10.5)	0.000	ns	
		Total path length of left instrument (cm)	102.5 (76.59 - 115.97)	94.1 (75.43 - 109.45)	0.238	ns	
		Total path length of right instrument (cm)	133.6 (78 - 166.41)	88.4 (75.94 - 216.46)	0.000	0.045	0.002
		Total path length (L+R)	241.9 (161.22 - 279.12)	187.8 (154.18 - 308.36)	0.000	ns	
		Total time	50.2 (44.33 - 60.17)	51.5 (35.01 - 69.13)	0.000	ns	
4	Clip Applying	Accuracy rate - applied clips (%)	82 (69.64 - 100)	88.8 (77.14 - 95)	0.000	ns	
		Average speed of left instrument movement (cm/sec)	2.8 (2.46 - 3.87)	3.1 (2.34 - 4.89)	0.616	ns	
		Average speed of right instrument movement (cm/sec)	3 (2.42 - 3.45)	3.3 (2.51 - 3.78)	0.279	ns	
		Economy of movement - left instrument (%)	31.2 (13.31 - 46.86)	40.6 (8.87 - 58.87)	0.218	ns	
		Economy of movement - right instrument (%)	56.2 (32.73 - 74.28)	63.8 (36.05 - 95.17)	0.062	ns	
		Ideal path length of left instrument (cm)	21.2 (8.55 - 64.6)	22.7 (13.89 - 34.73)	0.009	ns	
		Ideal path length of right instrument (cm)	73.7 (20.89 - 99.71)	74 (58.46 - 137.27)	0.961	ns	
		Number of clipped ducts	9 (9 - 9)	9 (9 - 9)	n/a	ns	
		Number of lost clips	2.3 (0 - 4)	1.3 (0.5 - 3)	0.000	ns	
		Number of movements of left instrument	40.8 (24.5 - 63.5)	29.8 (14 - 56)	0.000	ns	
		Number of movements of right instrument	51.5 (29.5 - 82)	46 (29.5 - 67.5)	0.145	ns	

		Total number of movements	97 (62.5 - 126.5)	67.3 (46.5 - 123.5)	0.000	ns	
		Relevant path length - left instrument (cm)	75.2 (50.53 - 157.55)	99.1 (55.38 - 237.13)	0.000	ns	
		Relevant path length - right instrument (cm)	123.5 (58.27 - 164.29)	137 (81.43 - 214.33)	0.201	ns	
		Total Number of clipping attempts	11.3 (9 - 13)	10.3 (9.5 - 12)	0.000	ns	
		Total path length of left instrument (cm)	104.6 (66.94 - 193.15)	102.8 (26.45 - 258.29)	0.000	ns	
		Total path length of right instrument (cm)	155.4 (94.85 - 192.28)	164.6 (104.87 - 259.71)	0.905	ns	
		Total path length (L+R)	273.1 (195.25 - 301.82)	255.8 (178.86 - 518)	0.000	ns	
		Total time	76.8 (60.87 - 109.08)	79.1 (59.01 - 103.11)	0.000	ns	
5	Clipping and Grasping	Accuracy rate - applied clips (%)	84.2 (64.29 - 100)	90.5 (85.91 - 100)	0.000	0.043	0.029
		Average speed of left instrument movement (cm/sec)	4.4 (2.69 - 4.77)	3.5 (3.08 - 5.13)	0.235	ns	
		Average speed of right instrument movement (cm/sec)	3.1 (2.1 - 3.75)	2.7 (2.24 - 3.76)	0.315	ns	
		Economy of movement - clipper (%)	45.5 (36.68 - 68.02)	62.9 (49.44 - 80.93)	0.000	0.004	0.000
		Economy of movement - left instrument (%)	43.8 (24.81 - 63.4)	48.2 (40.03 - 69.2)	0.056	ns	
		Economy of movement - right instrument (%)	45.5 (36.68 - 68.02)	62.9 (49.44 - 80.93)	0.000	0.004	0.000
		Economy of movement - grasper (%)	43.8 (24.81 - 63.4)	48.2 (40.03 - 69.2)	0.056	ns	
		Ideal path length of clipper (cm)	107.9 (84.95 - 180.62)	117.7 (96.78 - 162.14)	0.003	ns	
		Ideal path length of grasper (cm)	139.3 (107.61 - 165.17)	119.8 (107.65 - 158.83)	0.787	ns	
		Number of clipped ducts	9 (9 - 9)	9 (9 - 9)	n/a	ns	
		Number of lost clips	1.8 (0 - 5)	1 (0 - 1.5)	0.000	0.022	0.029
		Number of movements of left instrument	82 (46 - 132)	65 (34.5 - 81)	0.000	0.017	0.000
		Number of movements of right instrument	84.3 (54.5 - 135.5)	56 (34.5 - 80.5)	0.000	0.003	0.000
		Total number of movements	167.3 (105 - 251)	124 (69 - 147)	0.000	0.005	0.000
		Relevant path length - clipper(cm)	265.2 (174.82 - 344.84)	193.1 (123.32 - 310.26)	0.000	0.037	0.001

		Relevant path length - grasper (cm)	352 (190.96 - 493.19)	262.3 (156.86 - 320.72)	0.000	0.017	0.000
		Relevant path length - left instrument (cm)	352 (190.96 - 493.19)	262.3 (156.86 - 320.72)	0.000	0.017	0.000
		Relevant path length - right instrument (cm)	265.2 (174.82 - 344.84)	193.1 (123.32 - 310.26)	0.000	0.037	0.001
		Total Number of clipping attempts	10.8 (9 - 14)	10 (9 - 10.5)	0.000	0.022	0.029
		Total path length of clipper (cm)	273.5 (181.61 - 351.5)	199.9 (126.2 - 320.28)	0.000	0.045	0.002
		Total path length of grasper (cm)	356.5 (193.32 - 496.01)	267.5 (159.27 - 325.41)	0.000	0.017	0.000
		Total path length of left instrument (cm)	356.5 (193.32 - 496.01)	267.5 (159.27 - 325.41)	0.000	0.017	0.000
		Total path length of right instrument (cm)	273.5 (181.61 - 351.5)	199.9 (126.2 - 320.28)	0.000	0.045	0.002
		Total path length (L+R)	613.6 (374.93 - 808.25)	470 (285.47 - 645.69)	0.000	0.031	0.000
		Total time	133.9 (85.54 - 177.04)	100.4 (77.88 - 126.05)	0.000	0.014	0.000
6	Two Handed Manoeuvres	Average speed of left instrument movement (cm/sec)	3.3 (2.82 - 4.2)	3.2 (2.97 - 3.68)	0.048	ns	
		Average speed of right instrument movement (cm/sec)	3.4 (2.49 - 4.23)	3.4 (2.84 - 3.82)	0.012	ns	
		Economy of movement - left instrument (%)	30.7 (24.5 - 43)	36.3 (31.27 - 45.79)	0.000	ns	
		Economy of movement - right instrument (%)	36.4 (25.74 - 47.89)	33 (26.77 - 37.68)	0.594	ns	
		Ideal path length of left instrument (cm)	46.2 (20.66 - 76.56)	59 (52.55 - 83.95)	0.001	0.023	0.022
		Ideal path length of right instrument (cm)	68.9 (41.15 - 98.2)	56.6 (27.97 - 94.07)	0.000	ns	
		Number of exposed green balls that are collected	8 (7 - 9)	8.5 (7.5 - 8.5)	0.000	ns	
		Number of lost balls which miss the basket	1 (0 - 2)	0.5 (0.5 - 1.5)	0.000	ns	
		Number of movements of left instrument	70.8 (45 - 115.5)	68.3 (38 - 91)	0.003	ns	
		Number of movements of right instrument	78.8 (45.5 - 127)	61.3 (37.5 - 128)	0.000	ns	
		Total number of movements	147.3 (104.5 - 242.5)	134 (89.5 - 186)	0.000	ns	
		Relevant path length - left instrument (cm)	141.5 (46.78 - 279.64)	164.5 (120.48 - 268.77)	0.425	ns	
		Relevant path length - right instrument (cm)	195.5 (120.71 - 265.56)	235.7 (84.67 - 332.78)	0.001	ns	

		Total path length of left instrument (cm)	208.5 (139.54 - 328.04)	228.7 (139.8 - 324.93)	0.014	ns
		Total path length of right instrument (cm)	256.9 (151.06 - 379.14)	235.6 (127.53 - 365.61)	0.002	ns
		Total path length (L+R)	472.3 (337.54 - 707.17)	429.3 (346.38 - 690.54)	0.000	ns
		Total time	124.7 (87.16 - 161.89)	119.9 (96.61 - 172.63)	0.012	ns
7	Cutting	Accuracy rate - cuts without injury (%)	100 (98.39 - 100)	100 (100 - 100)	0.000	ns
		Average speed of left instrument movement (cm/sec)	2.9 (1.88 - 3.76)	2.6 (1.91 - 2.96)	0.095	ns
		Average speed of right instrument movement (cm/sec)	3.1 (1.62 - 3.7)	3.1 (2.76 - 4.84)	0.010	ns
		Number of cutting maneuvers performed without causing injury	28.3 (22.5 - 35)	27 (20 - 32)	0.027	ns
		Number of movements of left instrument	33.3 (21.5 - 61)	35.8 (16 - 56)	0.000	ns
		Number of movements of right instrument	103 (64.5 - 192)	93.8 (56 - 135.5)	0.000	ns
		Total number of movements	135.3 (93.5 - 253)	131.8 (72 - 191.5)	0.000	ns
		Number of retraction operations without overstretch injuries to tissue	1.3 (0.5 - 5)	1.3 (1 - 1.5)	0.000	ns
		Safe retraction - overstretch (%)	50 (18.75 - 90)	28.5 (6.28 - 100)	0.000	ns
		Total Number of cutting maneuvers	28.5 (22.5 - 35)	27 (20 - 32)	0.028	ns
		Total Number of retraction operations	3.8 (1.5 - 17.5)	8 (1 - 24.5)	0.000	ns
		Total path length of left instrument (cm)	90.9 (49.78 - 195.81)	106.6 (39.42 - 169.53)	0.000	ns
		Total path length of right instrument (cm)	237.5 (118.98 - 348.48)	215.9 (115.49 - 271.04)	0.000	ns
		Total path length (L+R)	326.7 (181.75 - 542.48)	294.9 (154.91 - 439.7)	0.000	ns
		Total time	106.8 (61.07 - 141.47)	99.6 (58.58 - 140.69)	0.017	ns
8	Electrocautery	Accuracy rate - highlighted bands (%)	100 (97.62 - 100)	100 (85.71 - 100)	0.000	ns
		Average speed of left instrument movement (cm/sec)	2.4 (2.09 - 2.92)	2.4 (1.96 - 3.11)	0.000	ns
		Average speed of right instrument movement (cm/sec)	2.1 (1.94 - 2.52)	2.3 (1.81 - 2.66)	0.000	ns

		Efficiency of cautery (%)	91.2 (74.37 - 94.39)	91.5 (84.7 - 95.95)	0.000	ns	
		Number of highlighted bands that were cut	21 (20.5 - 21)	21 (18 - 21)	0.000	ns	
		Number of movements of left instrument	112 (76 - 165.5)	96 (56.5 - 182.5)	0.000	ns	
		Number of movements of right instrument	140 (99.5 - 202.5)	153 (91.5 - 177.5)	0.000	ns	
		Total number of movements	248.8 (175.5 - 368)	251 (148 - 360)	0.000	ns	
		Number of non-highlighted bands that were cut	0 (0 - 0.5)	0 (0 - 3)	0.000	ns	
		The time cautery is applied without appropriate contact with bands	3.7 (2.03 - 7.45)	3.3 (2.12 - 5.73)	0.016	ns	
		Time cautery is applied on non-highlighted bands	4.6 (0.8 - 6.59)	6.1 (2.3 - 11.06)	0.005	0.037	0.436
		Total cautery time	40.6 (29 - 48.95)	39.1 (33.11 - 51.4)	0.000	ns	
		Total path length of left instrument (cm)	255.4 (171.1 - 385.09)	236.5 (108.93 - 485.39)	0.628	ns	
		Total path length of right instrument (cm)	288.4 (210.31 - 460.12)	343.8 (233.5 - 528.39)	0.004	ns	
		Total path length (L+R)	567.3 (381.42 - 845.21)	606.5 (342.7 - 1013.78)	0.000	ns	
		Total time	248.4 (204.53 - 331.22)	231.6 (143.53 - 353.81)	0.011	ns	
9	Translocation of Objects	Average Number of translocations per object	8.3 (5.5 - 11.25)	7.3 (3.17 - 12.67)	0.000	ns	
		Average speed of left instrument movement (cm/sec)	2.4 (2.22 - 3.07)	2.5 (2.28 - 3.04)	0.143	ns	
		Average speed of right instrument movement (cm/sec)	2.6 (1.99 - 3.35)	3.2 (2.3 - 3.62)	0.297	0.044	0.303
		Efficiency of translocations (%)	50.7 (38.12 - 77.91)	58.2 (31.93 - 92.86)	0.000	ns	
		Number of dropped objects	20 (9 - 28.5)	20.5 (9.5 - 41.5)	0.000	ns	
		Number of movements of left instrument	714.5 (450 - 1066)	487 (292 - 892)	0.009	0.014	0.000
		Number of movements of right instrument	661 (525 - 1075)	599.5 (374 - 1022)	0.003	ns	
		Total number of movements	1375.5 (981 - 2141)	1105.5 (666 - 1914)	0.000	0.045	0.000
		Number of objects	6 (6 - 6)	6 (6 - 6)	n/a	ns	
		Number of properly placed objects	6 (6 - 6)	6 (6 - 6)	n/a	ns	

	Number of translocations	49.5 (33 - 67.5)	44 (19 - 76)	0.000	ns	
	Total path length of left instrument (cm)	1614.2 (1227.61 - 2575.89)	1178.2 (915.69 - 2101.75)	0.000	0.025	0.000
	Total path length of right instrument (cm)	1621.7 (1067.24 - 2800.32)	1702.4 (1118.84 - 2484.45)	0.000	ns	
	Total path length (L+R)	3247 (2334.54 - 5294.19)	3001.3 (2034.53 - 4586.2)	0.000	ns	
	Total time	543.4 (332.69 - 663.5)	391.9 (196.34 - 686.91)	0.005	0.025	0.000

Table 10 – Performance metrics, construct validity and learning curve results for all 9 basic skills tasks. *Learning curves were not analysed if construct validity was not demonstrated. Results were considered significant where p < 0.05. ND = normal distribution, CV = construct validity, LC = learning curve.*

Task	Task description	Metric	Novices Median (range)	Experts Median (range)	ND? *P* value	CV? *P* value	LC? *P* value
1	Dissecting the Mesenteric Window	Economy of motion (left)	2.1 (1.71 - 2.96)	2.5 (2.08 - 3.35)	0.000	ns	
		Economy of motion (right)	2.2 (1.45 - 3.75)	3.1 (2.51 - 3.43)	0.000	0.007	0.080
		Idle time	51.9 (23.78 - 156.34)	15.7 (14.89 - 22.07)	0.000	0.000	0.000
		Injury to the Appendicular artery was controlled (number of occurrences)	0 (0 - 0)	0 (0 - 0)	n/a	n/a	
		Injury to the Appendicular artery was recorded (number of occurrences)	0 (0 - 1)	0 (0 - 0)	0.000	ns	
		Number of movements of left instrument	53.8 (28 - 209)	20.3 (12 - 35)	0.000	0.000	0.000
		Number of movements of right instrument	187.3 (77.5 - 509.5)	77.3 (40 - 112.5)	0.000	0.003	0.000
		Total number of movements	233.8 (105.5 - 718.5)	233.8 (105.5 - 718.5)	0.000	0.001	0.000
		Total path length of left instrument (cm)	68.3 (27.64 - 213.12)	25.8 (14.09 - 38.91)	0.000	0.001	0.000
		Total path length of right instrument (cm)	235.9 (132.93 - 612.66)	122.9 (78.61 - 198.88)	0.000	0.012	0.001
		Total path length (L+R)	299.8 (184.36 - 755.48)	299.8 (184.36 - 755.48)	0.000	0.010	0.000
		Total procedure time	198.1 (73.28 - 518.14)	54.2 (42.9 - 69.48)	0.000	0.000	0.000
2	Dissecting the Mesoappendix and Clipping the Artery	Economy of motion (left)	2.4 (1.51 - 2.89)	2.4 (2.03 - 3.1)	0.000	ns	
		Economy of motion (right)	2.2 (1.6 - 4.44)	2.7 (2.44 - 3.67)	0.000	0.007	ns
		Idle time	112 (72.54 - 190.88)	54 (28.15 - 61.61)	0.000	0.000	0.000
		Injury to the Appendicular artery was controlled (number of occurrences)	0.3 (0 - 1)	0 (0 - 0.5)	0.000	ns	
		Injury to the Appendicular artery was recorded (number of occurrences)	0.3 (0 - 1)	0 (0 - 1)	0.000	ns	
		Number of movements of left instrument	67.8 (35 - 167)	34.3 (16.5 - 39)	0.000	0.001	0.000
		Number of movements of right instrument	346 (242 - 566)	161.3 (98.5 - 265.5)	0.000	0.000	0.000
		Total number of movements	420.8 (284.5 - 733)	420.8 (284.5 - 733)	0.000	0.000	0.000
		Total path length of left instrument (cm)	54.9 (25.15 - 148.4)	29.4 (18.68 - 33.1)	0.000	0.012	0.006

		Total path length of right instrument (cm)	479.6 (293.42 - 930.81)	270.3 (165.95 - 542.92)	0.000	0.005	0.004
		Total path length (L+R)	535.4 (319.88 - 974.11)	535.4 (319.88 - 974.11)	0.000	0.004	0.004
		Total procedure time	299.6 (224.18 - 593.25)	161.9 (104.09 - 185.25)	0.000	0.000	0.000
3	Clipping the Artery and Ligating the Appendix Using a Ligating Loop	Economy of motion (left)	2.6 (2.16 - 3.05)	3.2 (2.68 - 3.67)	0.000	0.001	0.000
		Economy of motion (right)	2.5 (1.77 - 3.62)	2.9 (2.58 - 3.79)	0.000	0.016	0.000
		Idle time	260.1 (178.95 - 408.19)	150.1 (95.17 - 179.55)	0.000	0.000	0.000
		Injury to the Appendicular artery was controlled (number of occurrences)	0 (0 - 0.5)	0 (0 - 0)	0.000	ns	
		Injury to the Appendicular artery was recorded (number of occurrences)	0.3 (0 - 1.5)	0 (0 - 0)	0.000	0.042	ns
		Number of movements of left instrument	228.5 (142 - 464.5)	152 (103 - 192)	0.000	0.001	0.000
		Number of movements of right instrument	616.8 (448 - 1192.5)	400.5 (286 - 606)	0.000	0.003	0.000
		Total number of movements	851.8 (624 - 1657)	851.8 (624 - 1657)	0.000	0.002	0.000
		Total path length of left instrument (cm)	296.6 (213.43 - 821.83)	265.1 (214.27 - 355.05)	0.000	ns	
		Total path length of right instrument (cm)	996.9 (637.25 - 2521.09)	758.1 (526.66 - 1206.5)	0.000	ns	
		Total path length (L+R)	1253.4 (931.89 - 3342.91)	1253.4 (931.89 - 3342.91)	0.000	ns	
		Total procedure time	766.3 (493.56 - 1241.08)	459.2 (329.71 - 521.49)	0.000	0.000	0.000
4	Division of the Mesoappendix and Base of the Appendix Using a Stapler	Economy of motion (left)	2 (1.6 - 3.39)	2.4 (2.19 - 3.54)	0.000	0.007	ns
		Economy of motion (right)	2.7 (1.83 - 3.79)	3.5 (2.94 - 4.08)	0.016	0.010	0.000
		Idle time	40.7 (26.08 - 79.08)	25.9 (19.8 - 42.3)	0.000	0.004	0.000
		Number of movements of left instrument	78.3 (44.5 - 129.5)	57.3 (33 - 97)	0.000	ns	
		Number of movements of right instrument	189.5 (114 - 300)	107.8 (82 - 178)	0.000	0.012	0.000
		Total number of movements	261.5 (158.5 - 429.5)	261.5 (158.5 - 429.5)	0.000	0.016	0.000
		Total path length of left instrument (cm)	67 (43.78 - 131.48)	61.2 (36.7 - 191.42)	0.000	ns	

		Total path length of right instrument (cm)	293.6 (225.66 - 409.26)	253.7 (170.93 - 332.78)	0.000	ns	
		Total path length (L+R)	351.1 (275.43 - 540.74)	351.1 (275.43 - 540.74)	0.000	ns	
		Total procedure time	242.5 (168.58 - 459.51)	173.4 (134.96 - 215.96)	0.000	0.002	0.000
5	Control of the Artery Using Energy	Economy of motion (left)	2.9 (2.36 - 3.78)	3.3 (2.98 - 4.57)	0.004	0.025	0.000
		Economy of motion (right)	2.9 (2.3 - 4.02)	3.5 (2.95 - 4.56)	0.054	0.007	0.000
		Idle time	139.8 (100.94 - 193.08)	90.5 (38.96 - 109.12)	0.004	0.000	0.000
		Injury to the Appendicular artery was controlled (number of occurrences)	0 (0 - 0.5)	0 (0 - 0.5)	0.000	ns	
		Injury to the Appendicular artery was recorded (number of occurrences)	0 (0 - 0.5)	0 (0 - 0.5)	0.000	ns	
		Number of movements of left instrument	157.8 (109.5 - 267)	112.8 (79.5 - 137.5)	0.000	0.000	0.000
		Number of movements of right instrument	261.3 (223 - 459)	165.3 (149 - 239.5)	0.000	0.000	0.000
		Total number of movements	425.5 (363.5 - 726)	425.5 (363.5 - 726)	0.000	0.000	0.000
		Total path length of left instrument (cm)	266.4 (184.86 - 459.24)	239.9 (174.59 - 287.66)	0.000	ns	
		Total path length of right instrument (cm)	484.3 (387.51 - 761.77)	404 (313.86 - 490.46)	0.000	0.012	0.012
		Total path length (L+R)	750.9 (572.37 - 1172.48)	750.9 (572.37 - 1172.48)	0.000	0.016	0.002
		Total procedure time	409 (315.18 - 648.43)	269.5 (192.47 - 324.37)	0.000	0.000	0.000

Table 11 – Performance metrics, construct validity and learning curve results for the appendicectomy five procedural tasks. *Learning curves were not analysed if construct validity was not demonstrated. Results were considered significant where $p < 0.05$. ND = normal distribution, CV = construct validity, LC = learning curve. *The number of occurrences were 0 in both groups.*

Task	Task description	Metric	Novices Median (range)	Experts Median (range)	ND? *P* value	CV? *P* value	LC? *P* value
1	Appendectomy Complete Procedure	Economy of motion (left)	2.5 (2.21 - 3.41)	2.9 (2.42 - 4.08)	0.047	ns	
		Economy of motion (right)	2.1 (1.66 - 2.68)	2.8 (2.09 - 3.46)	0.200	0.006	0.001
		Idle time	271.1 (204.24 - 368.58)	140.1 (115.53 - 178.02)	0.000	0.000	0.000
		Injury to the Appendicular artery was controlled (number of occurrences)	0 (0 - 0.5)	0 (0 - 0.5)	0.000	ns	
		Injury to the Appendicular artery was recorded (number of occurrences)	0 (0 - 0.5)	0 (0 - 0.5)	0.000	ns	
		Number of movements of left instrument	284.5 (147.5 - 350)	120 (100 - 189.5)	0.000	0.000	0.000
		Number of movements of right instrument	611.5 (476.5 - 774.5)	313 (213 - 351.5)	0.000	0.000	0.000
		Total number of movements	909.5 (624 - 1124.5)	433 (313 - 541)	0.000	0.000	0.012
		Total path length of left instrument (cm)	366.1 (278.41 - 529.76)	245.8 (196.11 - 298.56)	0.000	0.000	0.001
		Total path length of right instrument (cm)	814.7 (672.65 - 1387.22)	574 (432.11 - 708.39)	0.000	0.000	0.001
		Total path length (L+R)	1231.5 (951.06 - 1855.17)	821.8 (628.22 - 944.61)	0.000	0.000	0.000
		Total procedure time	895.2 (653.99 - 986.49)	438.3 (351.16 - 539.89)	0.000	0.000	0.000

***Table 12 – Performance metrics, construct validity and learning curve results for the full appendicectomy procedure.** Learning curves were not analysed if construct validity was not demonstrated. Results were considered significant where $p < 0.05$. ND = normal distribution, CV = construct validity, LC = learning curve.*

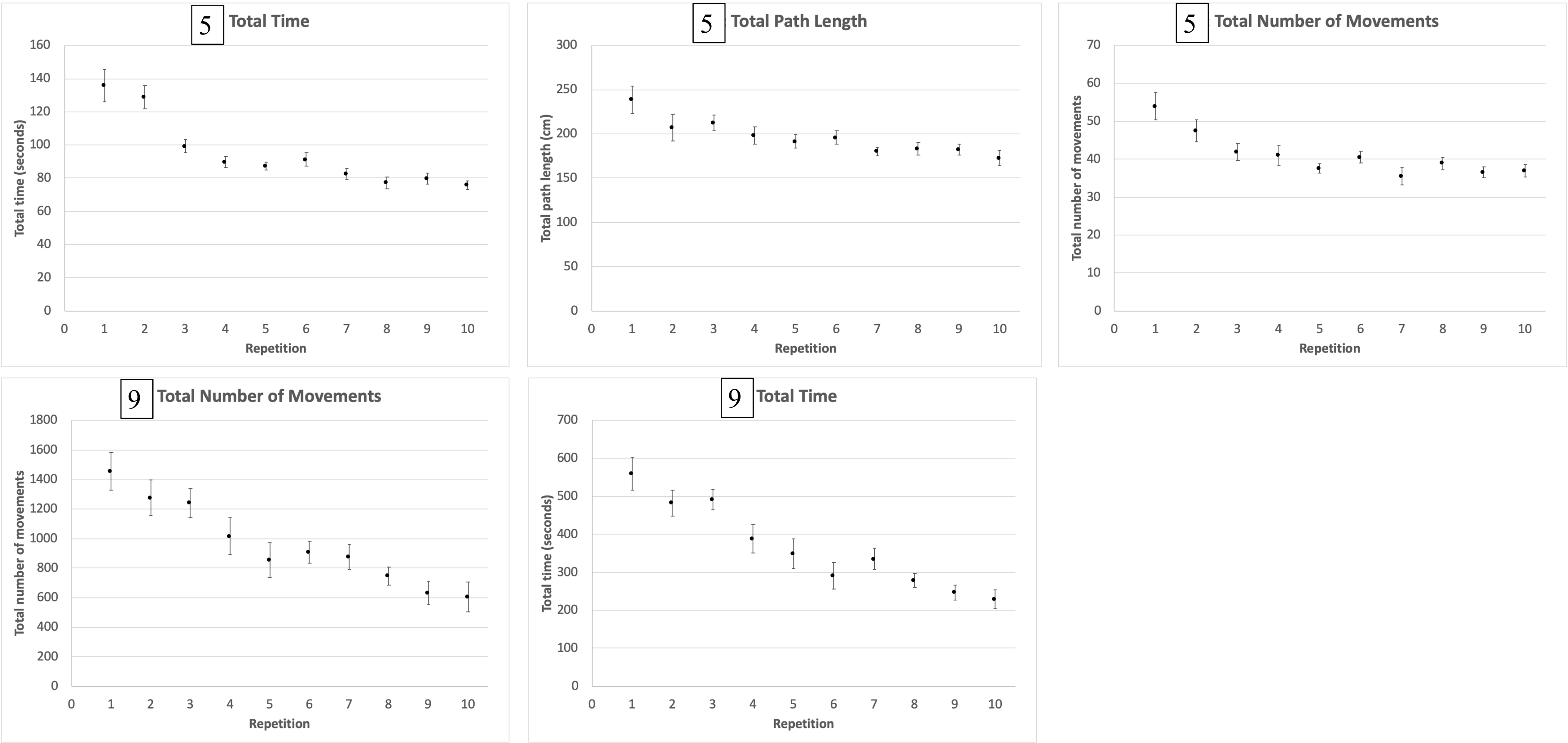
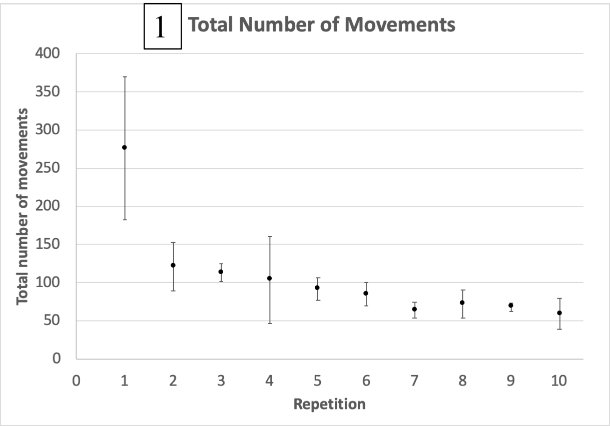
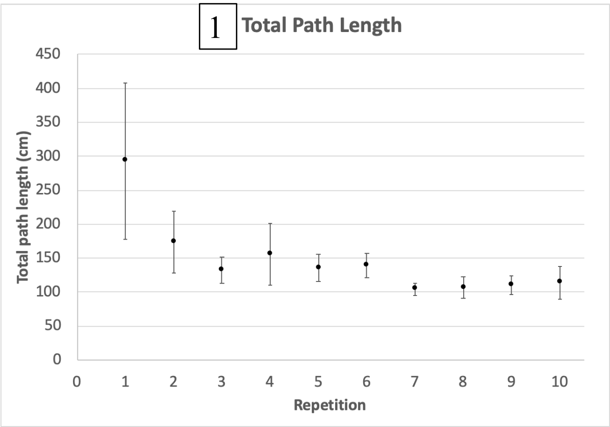
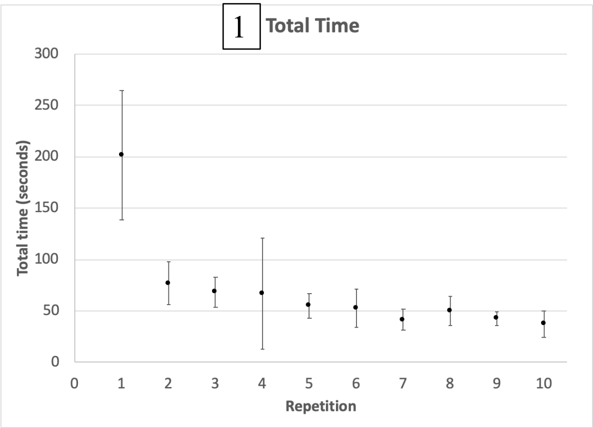
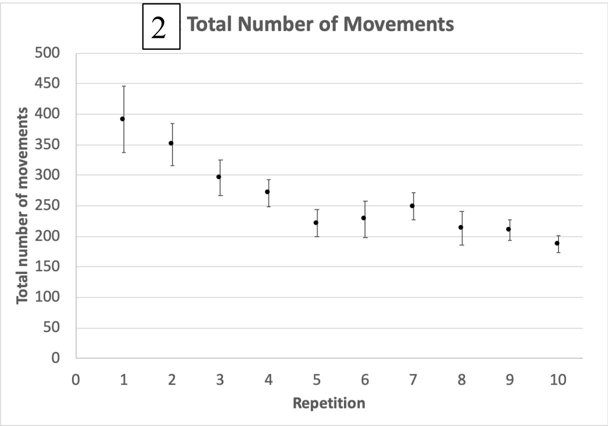


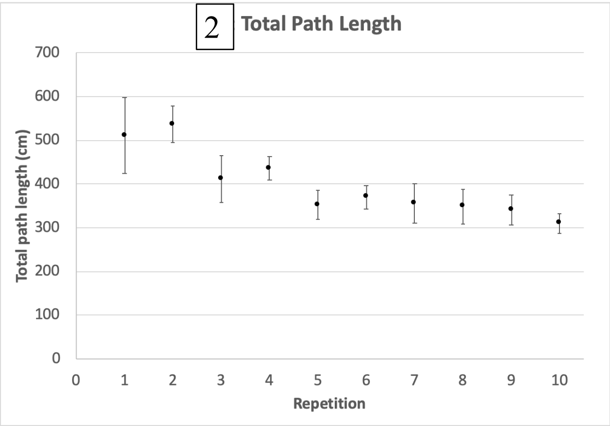
Figure 6 – Learning curves for novices' performance of selected metrics from Basic Skills Task 5 and Task 9 (Subgroup N1). *Median and error bars representing 95% confidence intervals are presented for novices' repetitions 1 through 10. Both construct validity and learning curves were statistically significant for all of these metrics (p = 0.000).*

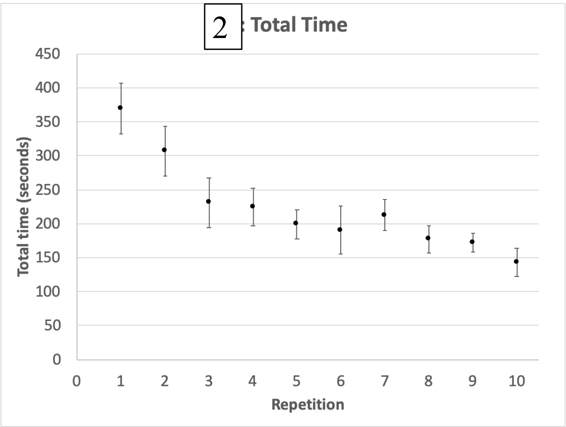

1
Total Number of Movements
Total number of movements
0
50
100
150
200
250
300
350
400
Repetition
0
1
2
3
4
5
6
7
8
9
10

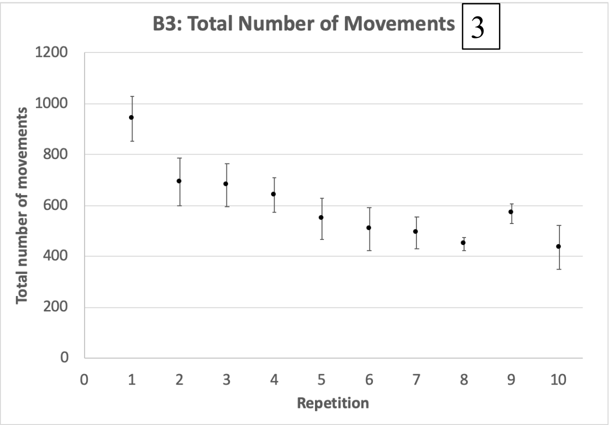

1
Total Path Length
Total path length (cm)
0
50
100
150
200
250
300
350
400
450
Repetition
0
1
2
3
4
5
6
7
8
9
10

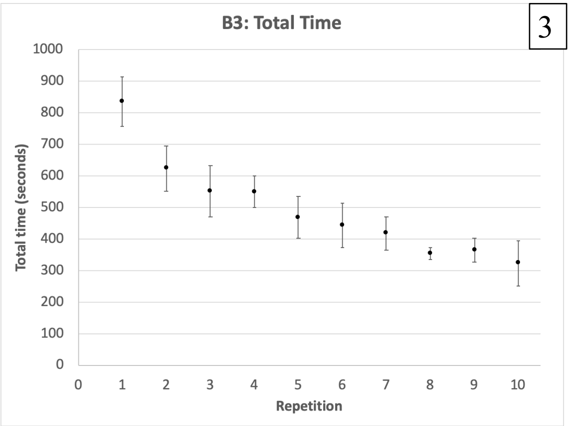

1
Total Time
Total time (seconds)
0
50
100
150
200
250
300
Repetition
0
1
2
3
4
5
6
7
8
9
10

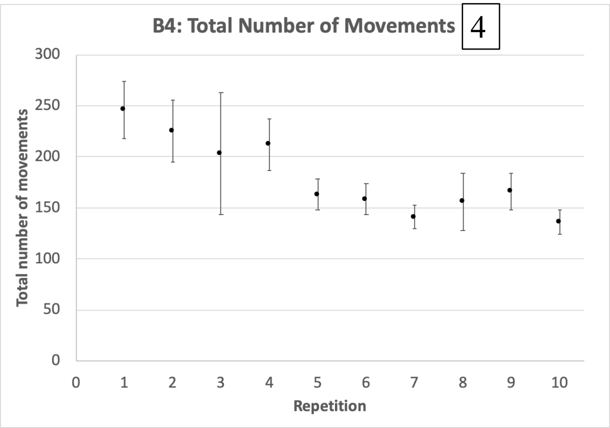

2
Total Number of Movements
Total number of movements
0
50
100
150
200
250
300
350
400
450
500
Repetition
0
1
2
3
4
5
6
7
8
9
10

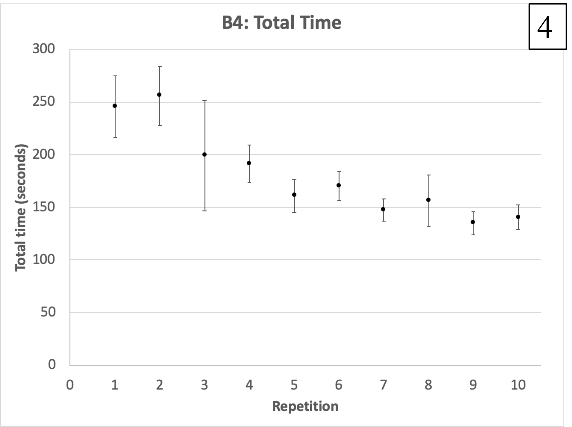

2
Total Path Length
Total path length (cm)
0
100
200
300
400
500
600
700
Repetition
0
1
2
3
4
5
6
7
8
9
10


2
Total Time
Total time (seconds)
0
50
100
150
200
250
300
350
400
450
Repetition
0
1
2
3
4
5
6
7
8
9
10


B3: Total Number of Movements
3
Total number of movements
0
200
400
600
800
1000
1200
0 1 2 3 4 5 6 7 8 9 10
Repetition


B3: Total Time
3
Total time (seconds)
0
100
200
300
400
500
600
700
800
900
1000
0 1 2 3 4 5 6 7 8 9 10
Repetition


B4: Total Number of Movements
4
Total number of movements
0
50
100
150
200
250
300
0 1 2 3 4 5 6 7 8 9 10
Repetition


B4: Total Time
4
Total time (seconds)
0
50
100
150
200
250
300
0 1 2 3 4 5 6 7 8 9 10
Repetition

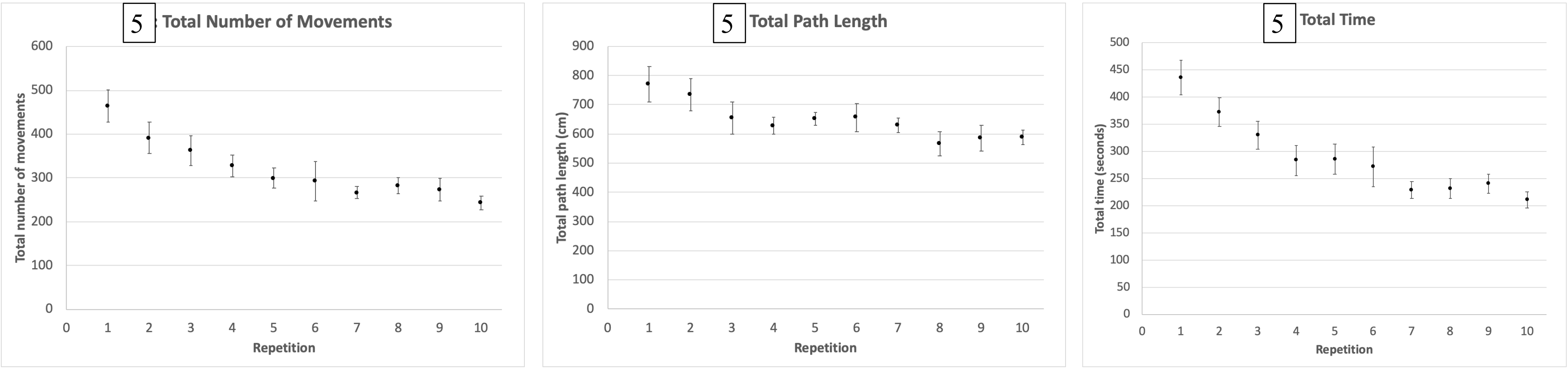


Figure 7 – Learning curves for novices' performance of selected metrics from the Procedural Tasks 1 to 5 (Subgroup N2). *Median and error bars representing 95% confidence intervals are presented for novices' repetitions 1 through 10 (apart from Procedural Task 5 Total Time, displaying mean and standard deviation due to data normality. Both construct validity and learning curves were significant for all of these metrics (p ≤ 0.005).*

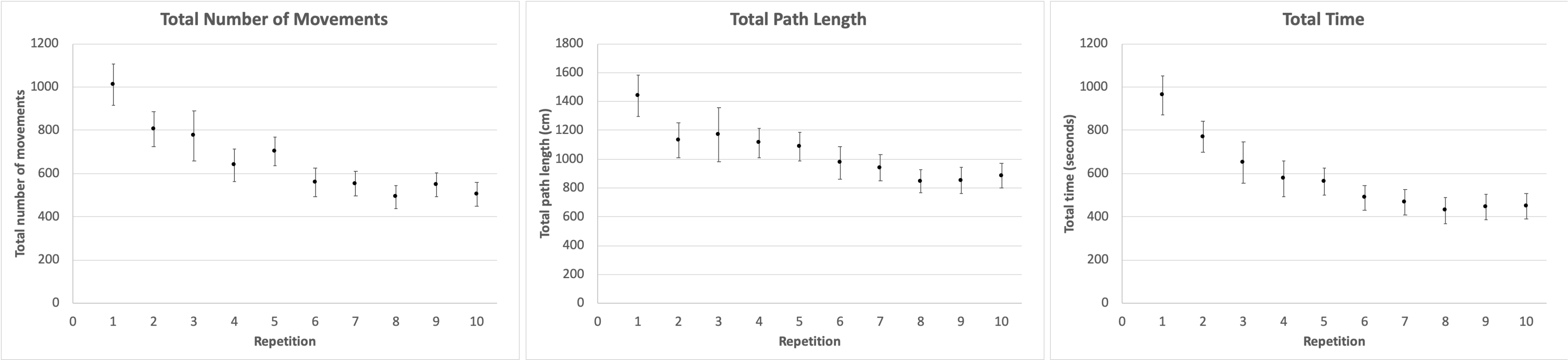


Figure 8 – Learning curves for novices' performance of selected metrics from the full laparoscopic appendicectomy procedure (Subgroup N3). *Median and error bars representing 95% confidence intervals are presented for novices' repetitions 1 through 10. Both construct validity and learning curves were statistically significant for all of these metrics ($p \leq 0.02$).*

3.1.3 Face Validity

Results from the face validity survey are presented in Table 13. Overall, novices and experts felt the appendicectomy module was moderately realistic, although haptic feedback was rated as only mildly realistic or neutral, and mildly realistic for the Endo-loop and Endo-catch devices. Participants strongly believed the module is useful for developing laparoscopic psychomotor skills, and moderately agreed that it is useful for the safe performance of a real appendicectomy. In particular, participants strongly felt that participation was enjoyable and they moderately (novices) and strongly (experts) agreed that mandatory completion should be considered prior to performing laparoscopic surgery on real patients.

Regarding the LAP Mentor VR environment, please rate the following with a number between 1 and 7:
(1 = not realistic, 4 = neutral, 7 = identical to reality)

	Novices	**Experts**
Realism in general:	5.14	5.06
Camera-navigation:	5.00	5.65
Instruments - Clip applier:	5.73	5.59
Instrument - Grasper:	5.73	5.35
Instruments - Maryland dissector (for blunt dissection):	5.45	5.24
Instruments - Marlyand dissector (for diathermy):	5.59	5.41
Instruments - Scissors:	5.32	5.59
Instruments - Endo-loop ligature:	4.36	4.06
Instruments - Endo-catch bag:	4.59	4.94
Instruments - Endo-stapler:*	5.36	4.88
Instruments - Harmonic device:*	5.18	5.38
Tissue reactions to manipulation:	5.00	4.53
Force ('haptic') feedback:	4.82	4.00
Appearance of the organs:	4.86	5.24

Please rate the usefulness of the LAP Mentor VR appendicectomy module in developing the following skills by using a number between 1 and 7:
(1 = useless, 4 = neutral, 7 = very useful)

	Novices	**Experts**
Hand-eye co-ordination:	6.41	6.29
Camera navigation:	4.77	5.18
Depth perception:	6.05	5.76
Safe performance of real laparoscopic appendicectomy:	5.86	5.24

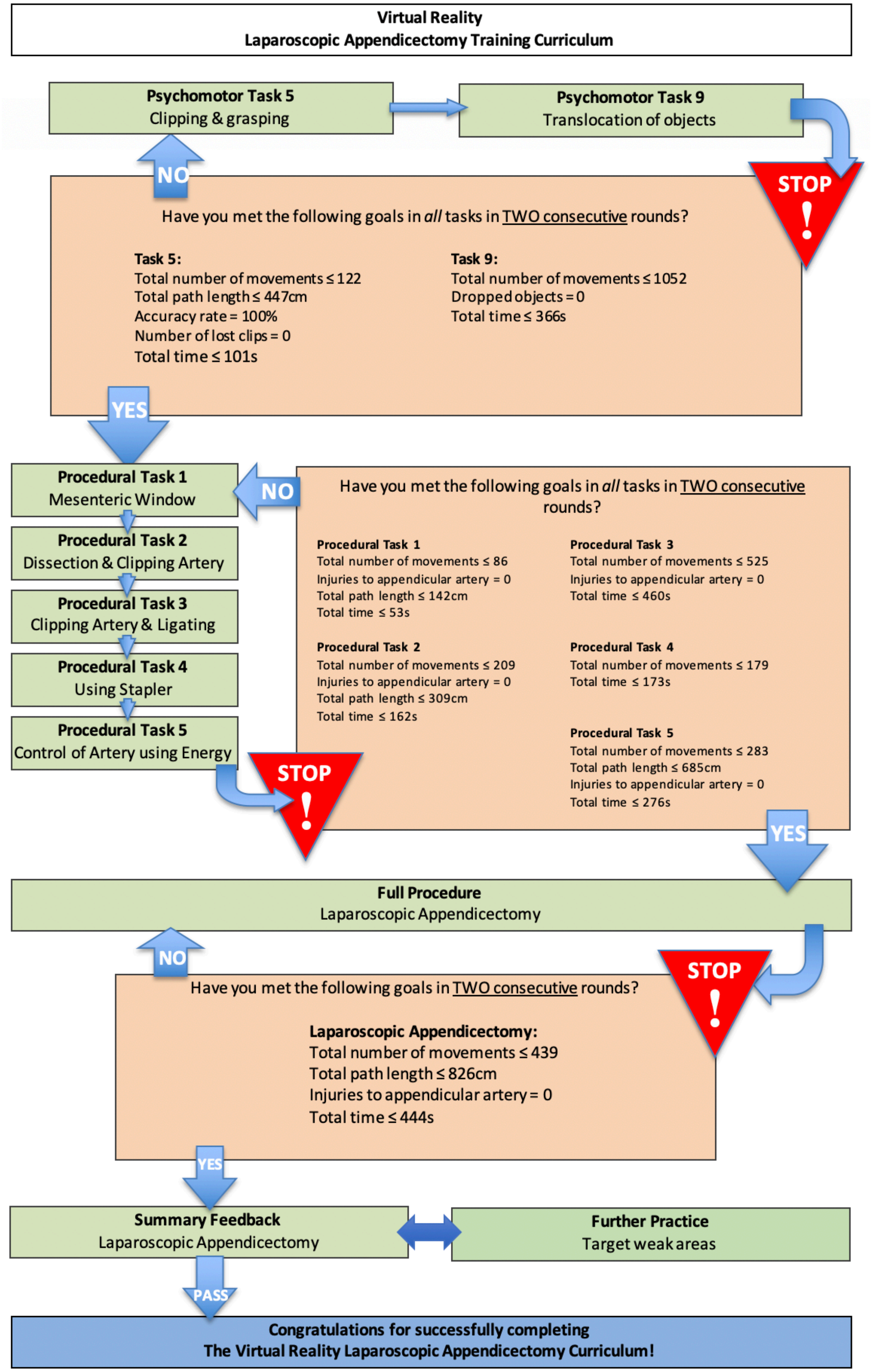
Regarding the LAP Mentor VR appendicectomy module, please rate the following statements with a number between 1 and 7:
(4 = neutral):

	Novices	**Experts**
The LAP Mentor VR appendicectomy sessions were: *(1 = boring, 7 = enjoyable)*	6.68	6.71
Before being allowed to perform laparoscopic surgery on real patients, a course based on the LAP Mentor VR appendicectomy module should be: *(1 = optional, 7 = mandatory)*	5.73	6.41

Table 13 – Face validity for the LAP Mentor appendicectomy module. *VR = virtual reality. Survey results from appendicectomy procedural tasks and full appendicectomy procedure were combined (N2 with N3, and E2 with E3): n = 39 (22 novices, 17 experts). * = these instruments were rated by participants of groups N2 and E2 only as they were not used by groups N3 and E3 (n = 11 novices and 8 experts).*

3.2 Part 2 – Construction of the VR LA Curriculum

The VR LA curriculum is presented in Figure 9. The rationale for inclusion or exclusion of each metric (according to the criteria outlined in the methodology) across the BS tasks, the LA procedural tasks and FP is presented in Table 14, Table ***15*** and Table ***16***, respectively. Benchmark scores were set at the 50th percentile of expert performances for all included metrics. Experts' 50th, 75th and 90th percentile scores for each metric included in the curriculum is presented in Table 17.



Virtual Reality
Laparoscopic Appendicectomy Training Curriculum

Psychomotor Task 5
Clipping & grasping

Psychomotor Task 9
Translocation of objects

Have you met the following goals in *all* tasks in TWO consecutive rounds?

Task 5:
Total number of movements ≤ 122
Total path length ≤ 447cm
Accuracy rate = 100%
Number of lost clips = 0
Total time ≤ 101s

Task 9:
Total number of movements ≤ 1052
Dropped objects = 0
Total time ≤ 366s

Procedural Task 1
Mesenteric Window

Procedural Task 2
Dissection & Clipping Artery

Procedural Task 3
Clipping Artery & Ligating

Procedural Task 4
Using Stapler

Procedural Task 5
Control of Artery using Energy

Have you met the following goals in *all* tasks in TWO consecutive rounds?

Procedural Task 1
Total number of movements ≤ 86
Injuries to appendicular artery = 0
Total path length ≤ 142cm
Total time ≤ 53s

Procedural Task 2
Total number of movements ≤ 209
Injuries to appendicular artery = 0
Total path length ≤ 309cm
Total time ≤ 162s

Procedural Task 3
Total number of movements ≤ 525
Injuries to appendicular artery = 0
Total time ≤ 460s

Procedural Task 4
Total number of movements ≤ 179
Total time ≤ 173s

Procedural Task 5
Total number of movements ≤ 283
Total path length ≤ 685cm
Injuries to appendicular artery = 0
Total time ≤ 276s

Full Procedure
Laparoscopic Appendicectomy

Have you met the following goals in TWO consecutive rounds?

Laparoscopic Appendicectomy:
Total number of movements ≤ 439
Total path length ≤ 826cm
Injuries to appendicular artery = 0
Total time ≤ 444s

Summary Feedback
Laparoscopic Appendicectomy

Further Practice
Target weak areas

Congratulations for successfully completing
The Virtual Reality Laparoscopic Appendicectomy Curriculum!

Figure 9 – Curriculum for laparoscopic appendicectomy on the LAP Mentor. *The median of experts' scores from the construct validity assessment were set as the proficiency criteria for progression through the curriculum. Participants following the curriculum may practice for up to 45 min per session, with no more than one session per day. Progression occurs when criteria have been met only at the start of the session to ensure consistency and retention of skill from the previous session, and to guarantee a greater number of spaced sessions. Those unable to progress after five sessions must request an instructor feedback session. Curriculum appeared in The American Journal of Surgery (2020)* (104).

Task	Metric	Description	Include?	Comments
3	Eye-Hand Coordination	Economy of movement - right instrument (%)	No	Although CV demonstrated, too narrow and isolated to warrant the task’s inclusion.
		Relevant path length - right instrument (cm)	No	Unclear how to affect improvement.
		Total path length of right instrument (cm)	No	This would discriminate according to participant dextrality.
5	Clipping and Grasping	Accuracy rate - applied clips (%)	Yes	To avoid rewarding speed at the expense of error.
		Economy of movement - clipper (%)	No	Whilst important, it is a product of task completion time and number of movements, which are both easier metrics to comprehend.
		Economy of movement - right instrument (%)	No	
		Number of lost clips	Yes	To avoid rewarding speed at the expense of error.
		Number of movements of left instrument	No	Implies confident, purposeful, and more economic movement relying less on ‘trial and error’ as development progresses. However, use total movements to counter for dextrality.
		Number of movements of right instrument	No	
		Total number of movements	Yes	To encourage only purposeful rather than superfluous movements - important to set these as benchmarks as this forms part of the foundation of laparoscopic psychomotor skill.
		Relevant path length - clipper(cm)	No	Unclear how to affect improvement.
		Relevant path length - grasper (cm)	No	
		Relevant path length - left instrument (cm)	No	
		Relevant path length - right instrument (cm)	No	
		Total Number of clipping attempts	No	It has been demonstrated that a clipping attempt may result in software failure where the clip doesn't deploy, and ‘number of lost clips’ is a more reliable measure of accuracy. Path length is an important marker of economy of movement. Use total path length to counter for dextrality (not specifying instrument type makes this simpler to comprehend).
		Total path length of clipper (cm)	No	
		Total path length of grasper (cm)	No	
		Total path length of left instrument (cm)	No	Use total path length to avoid confusion related to dextrality.
		Total path length of right instrument (cm)	No	
		Total path length (L+R)	Yes	A marker of economical movement that discourages direct and purposeful movements only, in order to achieve the shortest path length.
		Total time	Yes	An important metric that is easy to understand.

6	Two Handed Manoeuvres	Ideal path length of left instrument (cm)	No	Unclear what this means and how to affect improvement.
8	Electrocautery	Time cautery is applied on non-highlighted bands	No	The task is not useful where there is only a single construct-valid metric that is highly-specific.
9	Translocation of Objects	Average speed of right instrument movement (cm/sec)	No	Speed of instrument has potential to encourage unsafe habits.
		Number of dropped objects	Yes	To avoid both 'cheating' (manipulating object by dragging it around on the 'floor') and the reward of speed at the expense of errors.
		Number of movements of left instrument	No	This would unfairly discriminate with utility dependent upon the dextrality of the learner.
		Total number of movements	Yes	This is affected by economy of movement and the ability to mentally rehearse the effects upon an object of specific instrument movements (prior to moving them), and so are highly relevant laparoscopic psychomotor skills.
		Total path length of left instrument (cm)	No	Whilst this is useful, this task is long and difficult, and so it is felt to be too burdensome to be included alongside 'total number of movements.'
		Total time	Yes	An important metric that is easy to understand, and clearly a marker of efficient completion of this task.

Table 14 - Decision analysis for the inclusion of tasks and metrics from the Basic Skills module.

Task	Task description	Metric	Include?	Comments
1	Dissecting the Mesenteric Window	Idle time	No	Total time would be more relevant.
		Injury to the Appendicular artery was recorded (number of occurrences)	Yes	To avoid rewarding speed at the expense of error.
		Number of movements of left instrument	No	Implies confident, purposeful, and more economic movement relying less on ‘trial and error’ as development progresses. However, use total movements to counter for dextrality.
		Number of movements of right instrument	No	
		Total number of movements	Yes	
		Total path length of left instrument (cm)	No	Use total path length to avoid confusion related to dextrality.
		Total path length of right instrument (cm)	No	
		Total path length (L+R)	Yes	A marker of economical movement that discourages direct and purposeful movements only, in order to achieve the shortest path length.
		Total procedure time	Yes	Easy to understand, and a direct marker of efficiency of task completion.
2	Dissecting the Mesoappendix and Clipping the Artery	Idle time	No	Total time would be more relevant.
		Injury to the Appendicular artery was recorded (number of occurrences)	Yes	To avoid rewarding speed at the expense of error.
		Number of movements of left instrument	No	Implies confident, purposeful, and more economic movement relying less on ‘trial and error’ as development progresses. However, use total movements to counter for dextrality.
		Number of movements of right instrument	No	
		Total number of movements	Yes	
		Total path length of left instrument (cm)	No	Use total path length to avoid confusion related to dextrality.
		Total path length of right instrument (cm)	No	
		Total path length (L+R)	Yes	A marker of economical movement that discourages direct and purposeful movements only, in order to achieve the shortest path length.
		Total procedure time	Yes	Easy to understand, and a direct marker of efficiency of task completion.
3	Clipping the Artery and Ligating the Appendix Using a Ligating Loop	Economy of motion (left)	No	Whilst important, economy of motion is a product of task completion time and number of movements, which are both easier metrics to comprehend.
		Economy of motion (right)	No	
		Idle time	No	Total time would be more relevant.

		Injury to the Appendicular artery was recorded (number of occurrences)	Yes	To avoid rewarding speed at the expense of error.
		Number of movements of left instrument	No	Implies confident, purposeful, and more economic movement relying less on ‘trial and error’ as development progresses. However, use total movements to counter for dextrality.
		Number of movements of right instrument	No	
		Total number of movements	Yes	
		Total procedure time	Yes	Easy to understand, and a direct marker of efficiency of task completion.
4	Division of the Mesoappendix and Base of the Appendix Using a Stapler	Economy of motion (right)	No	This is a product of task completion time and number of movements, which are both easier metrics to comprehend. Also, isolated to right hand only, so raises confusion with regards to dextrality.
		Idle time	No	Total time would be more relevant.
		Number of movements of right instrument	No	Implies confident, purposeful, and more economic movement relying less on ‘trial and error’ as development progresses. However, use total movements to counter for dextrality.
		Total number of movements	Yes	
		Total procedure time	Yes	Easy to understand, and a direct marker of efficiency of task completion.
5	Control of the Artery Using Energy	Economy of motion (left)	No	Whilst important, economy of motion is a product of task completion time and number of movements, which are both easier metrics to comprehend.
		Economy of motion (right)	No	
		Idle time	No	Total time would be more relevant.
		Injury to the Appendicular artery was recorded (number of occurrences)	Yes	To avoid rewarding speed at the expense of error.
		Number of movements of left instrument	No	Implies confident, purposeful, and more economic movement relying less on ‘trial and error’ as development progresses. However, use total movements to counter for dextrality.
		Number of movements of right instrument	No	
		Total number of movements	Yes	
		Total path length of right instrument (cm)	No	A marker of economical movement that discourages direct and purposeful movements only, in order to achieve the shortest path length. Use total path length to avoid confusion related to dextrality.
		Total path length (L+R)	Yes	
		Total procedure time	Yes	Easy to understand, and a direct marker of efficiency of task completion.

Table 15 – Decision analysis for the inclusion of tasks and metrics from the procedural tasks of the Laparoscopic Appendicectomy module.

Task	Task description	Metric	Include?	Comments
1	Appendectomy Complete Procedure	Economy of motion (right)	No	Made redundant by more easily interpretable total movement and total path length metrics.
		Idle time	No	Total time would be more relevant.
		Injury to the Appendicular artery was controlled (number of occurrences)	Yes	To avoid rewarding speed at the expense of error.
		Number of movements of left instrument	No	Implies confident, purposeful, and more economic movement relying less on ‘trial and error’ as development progresses. However, use total movements to counter for dextrality.
		Number of movements of right instrument	No	
		Total number of movements	Yes	
		Total path length of left instrument (cm)	No	A marker of economical movement that discourages direct and purposeful movements only, in order to achieve the shortest path length. Use total path length to avoid confusion related to dextrality.
		Total path length of right instrument (cm)	No	
		Total path length (L+R)	Yes	
		Total procedure time	Yes	Easy to understand, and a direct marker of efficiency of task completion.

Table 16 – Decision analysis for the inclusion of metrics from the full procedure of the Laparoscopic Appendicectomy module.

Group/Task	Metric	Description	50th Percentile	75th Percentile	90th Percentile
A/5	Clipping and Grasping	Accuracy rate - applied clips (%)	100	100	100
		Number of lost clips	0	0	0
		Total number of movements	122	99	70
		Total path length (L+R)	447	382	297
		Total time	101	87	77
A/9	Translocation of Objects	Number of dropped objects	18	12	10
		Total number of movements	1052	780	600
		Total time	366	258	215
B/1	Dissecting the Mesenteric Window	Injury to the Appendicular artery was recorded (number of occurrences)	0	0	0
		Total number of movements	86	72	49
		Total path length (L+R)	142	106	97
		Total procedure time	53	46	35
B/2	Dissecting the Mesoappendix and Clipping the Artery	Injury to the Appendicular artery was recorded (number of occurrences)	0	0	0
		Total number of movements	209	149	133
		Total path length (L+R)	309	268	208
		Total procedure time	162	136	93
B/3	Clipping the Artery and Ligating the Appendix Using a Ligating Loop	Injury to the Appendicular artery was recorded (number of occurrences)	0	0	0
		Total number of movements	525	468	388
		Total procedure time	460	391	330
B/4	Division of the Mesoappendix and Base of the Appendix Using a Stapler	Total number of movements	179	130	112
		Total procedure time	173	154	136

B/5	Control of the Artery Using Energy	Injury to the Appendicular artery was recorded (number of occurrences)	0	0	0
		Total number of movements	283	246	221
		Total path length (L+R)	685	584	533
		Total procedure time	276	224	175
C/1	Appendectomy Complete Procedure	Injury to the Appendicular artery was controlled (number of occurrences)	0	0	0
		Total number of movements	439	396	324
		Total path length (L+R)	826	722	644
		Total procedure time	444	387	346

Table 17 – Expert performance benchmarks for all included metrics in the VR laparoscopic appendicectomy curriculum.

3.3 Part 3 – Implementation of the VR LA Curriculum

Prior to commencement of Part 3, the simulators in the CSBS had been upgraded from the LAP Mentor II to the LAP Mentor III. A notable difference is an updated delivery of haptic feedback. Seven participants completed the VR LA curriculum (Table 18) and three dropped out. Of the latter three, one dropped out following eight practice sessions and one feedback session without completing the first stage of the curriculum; the other two participants did not return after 9 and 7 practice sessions of the second stage, respectively. None of the dropouts provided a reason despite having been sent an email requesting this feedback and none provided a participant diary. One participant was declared as having successfully completed the curriculum despite only recording a single repetition of the final task (participant 7), and the reason for this discrepancy is unknown.

	Total
N	7
Male (female)	5 (2)
Participants with previous experience on VR simulator (mean, mins)	1 (17)
Participants with previous experience on video ('box') trainer	5
Previous full laparoscopic procedures as primary operator	0
Previous experience as camera navigator	6
RH dominance (LH)	7 (0)
Has played musical instrument	1
Has played video games	7

Table 18 – Baseline characteristic of those who successfully completed the curriculum. *Dropouts were not included.*

There was a total of 228 practice sessions completed in those that completed the curriculum (median 30 per participant, range 23 – 45), excluding feedback sessions. 170 repetitions were excluded from the analysis after having been positively identified as either resulting from system crashes or voluntary restarts due to software errors. After these exclusions, there was a total of 2,564 repetitions recorded across all tasks (median 401 per participant, range 179 – 423). PT1 was repeated the most by participants (median 72, range 21 – 153) and the full procedure the least (median 2, range 1 – 7). The total number of repetitions for each task, by participant, is presented in Figure 10 and Table 19. The learning curves for all metrics (apart from error scores) are presented in Figure 11.

The median number of feedback sessions per participant was 9 (range 7 – 10), which included those that were mandated to occur at the start of each stage of the curriculum and the final assessment. All 7 participants scored either 4 or 5 in all components of the final assessment using the modified GOALS assessment tool (Table 20). Individualised feedback from the final assessments is displayed in Table 21.

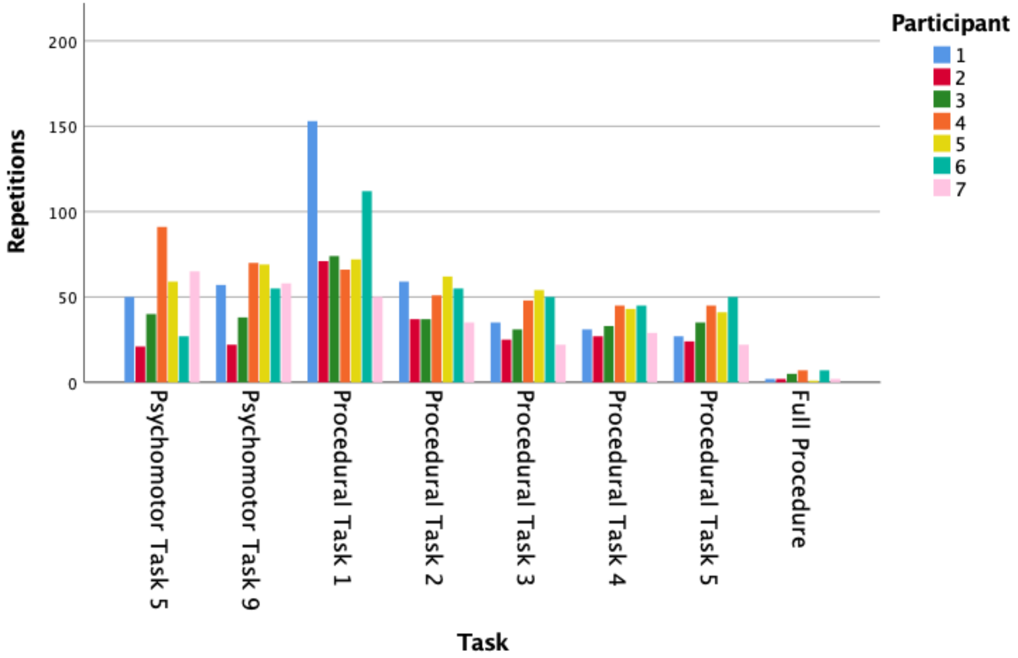
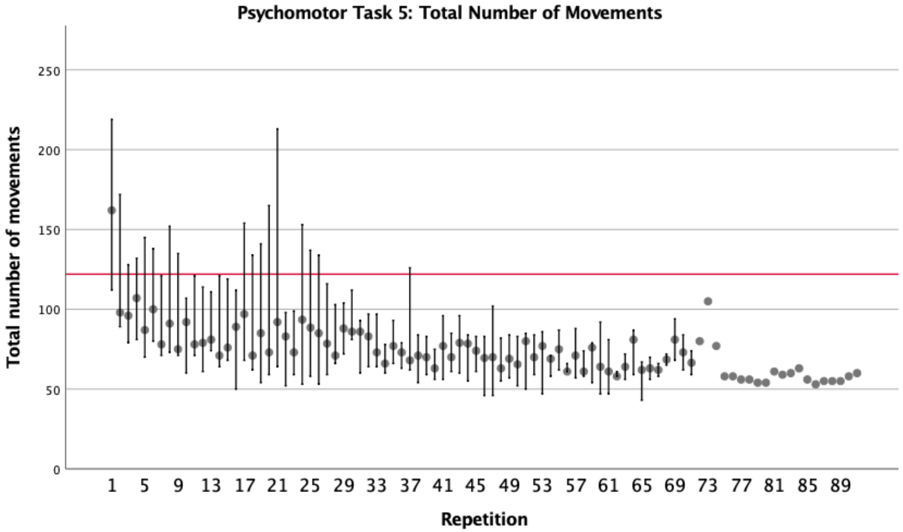
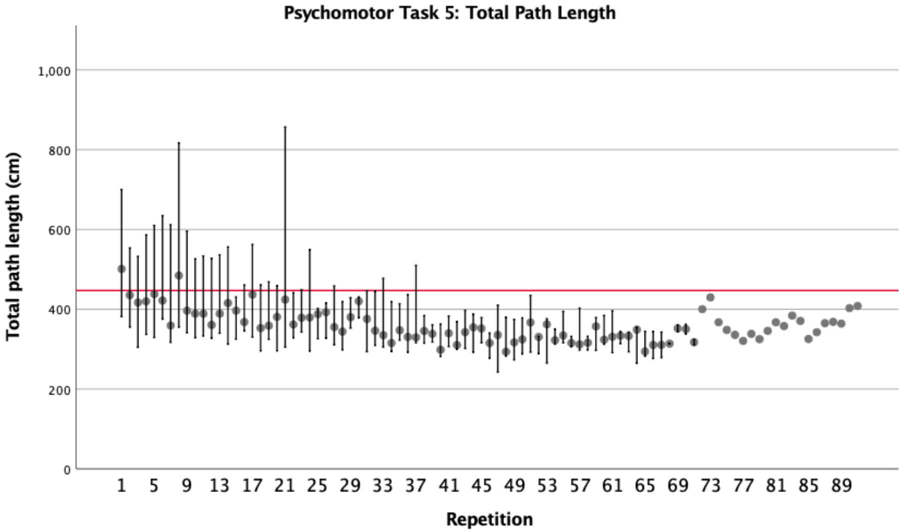
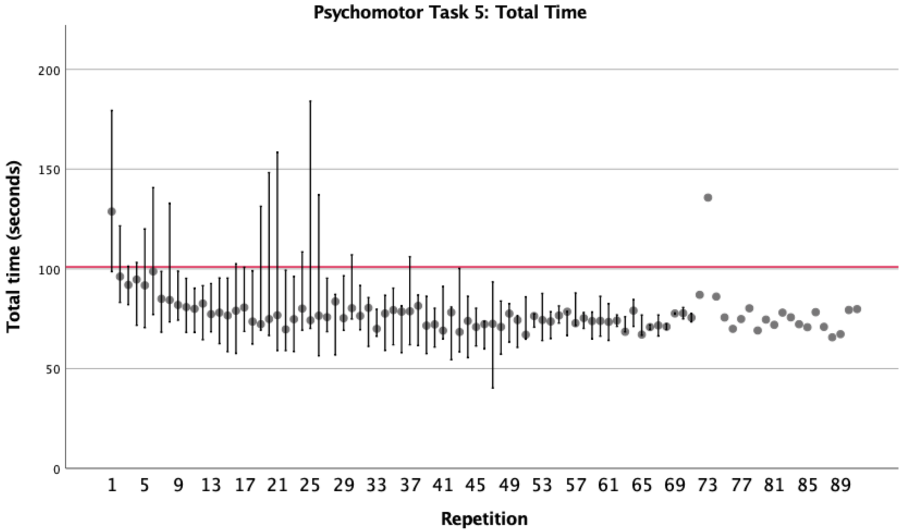
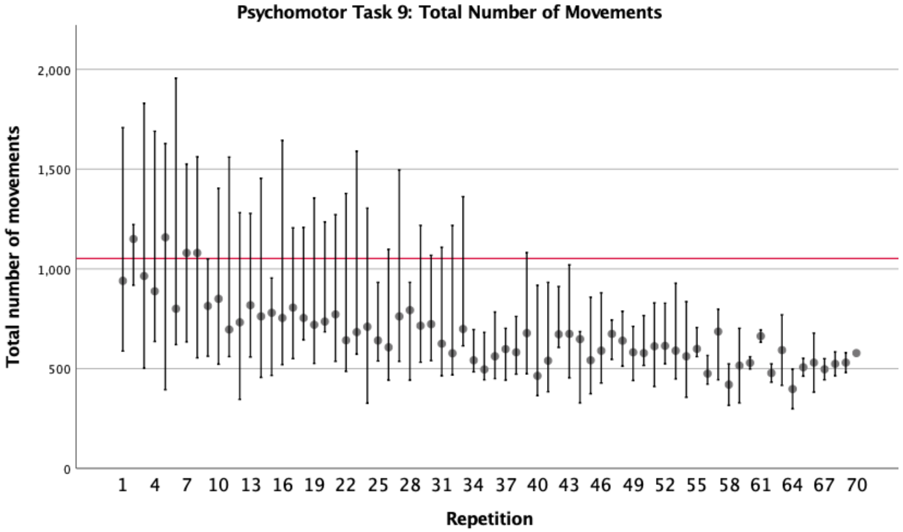


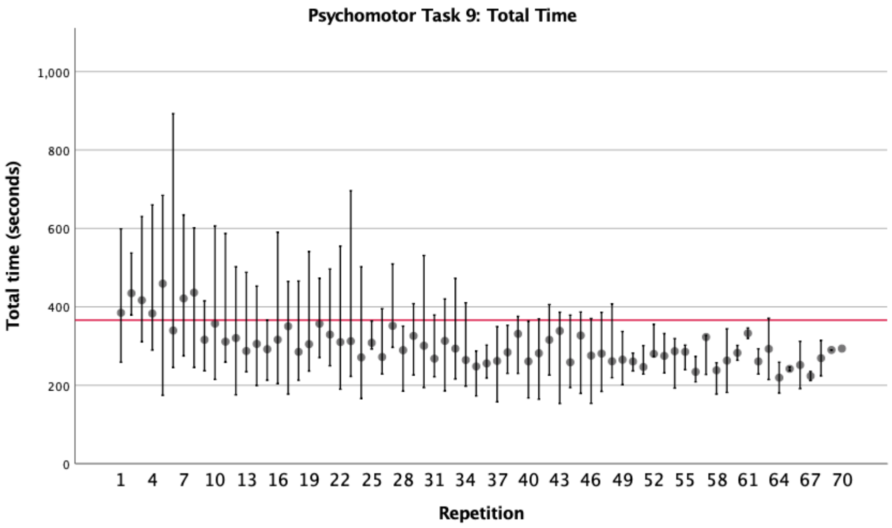
Figure 10 – Total number of repetitions for each task of the Laparoscopic Appendicectomy curriculum, by participant.

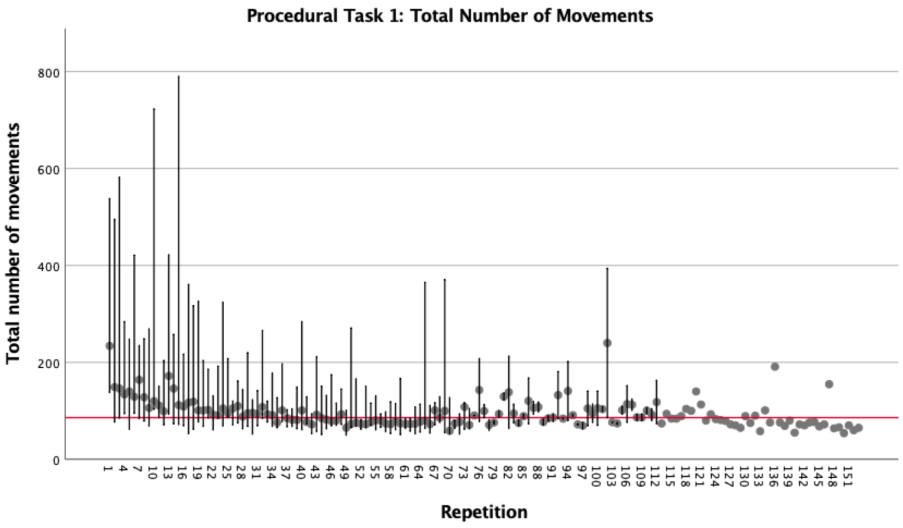

Psychomotor Task 5: Total Number of Movements
Total number of movements
0
50
100
150
200
250
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89
Repetition

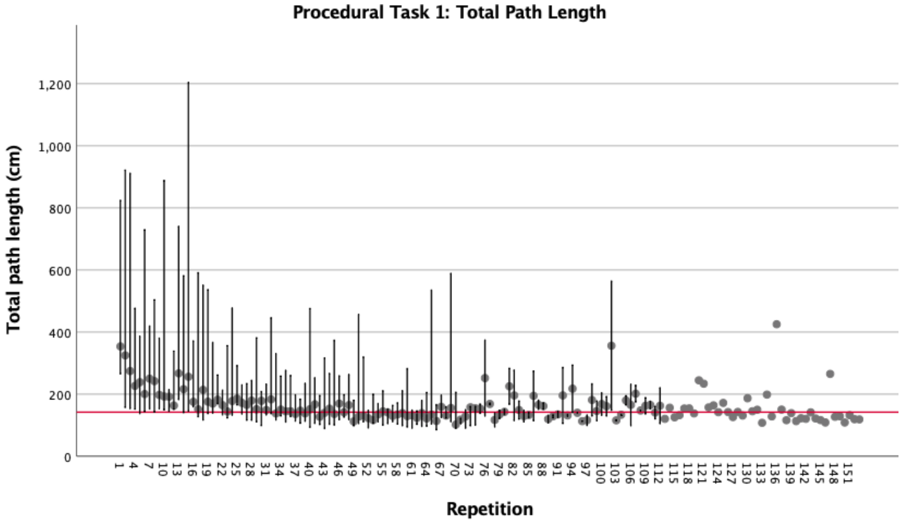

Psychomotor Task 5: Total Path Length
Total path length (cm)
0
200
400
600
800
1,000
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89
Repetition

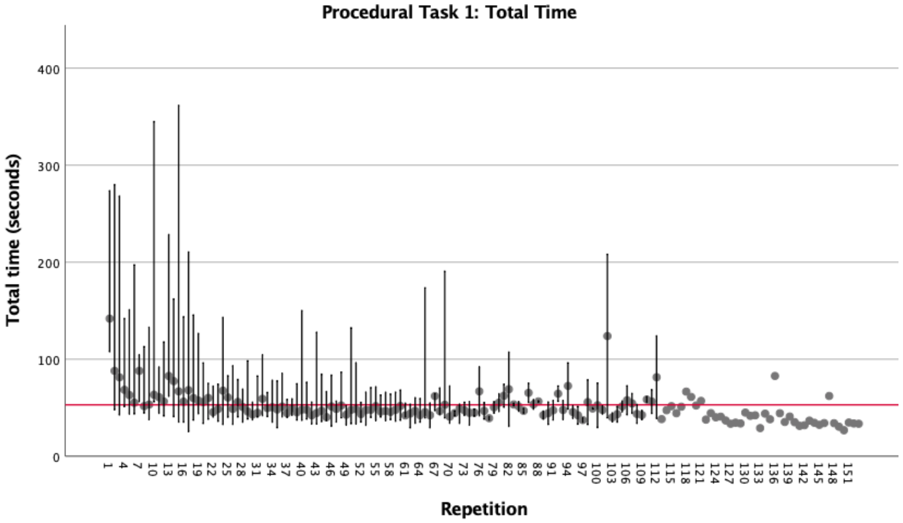

Psychomotor Task 5: Total Time
Total time (seconds)
0
50
100
150
200
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89
Repetition

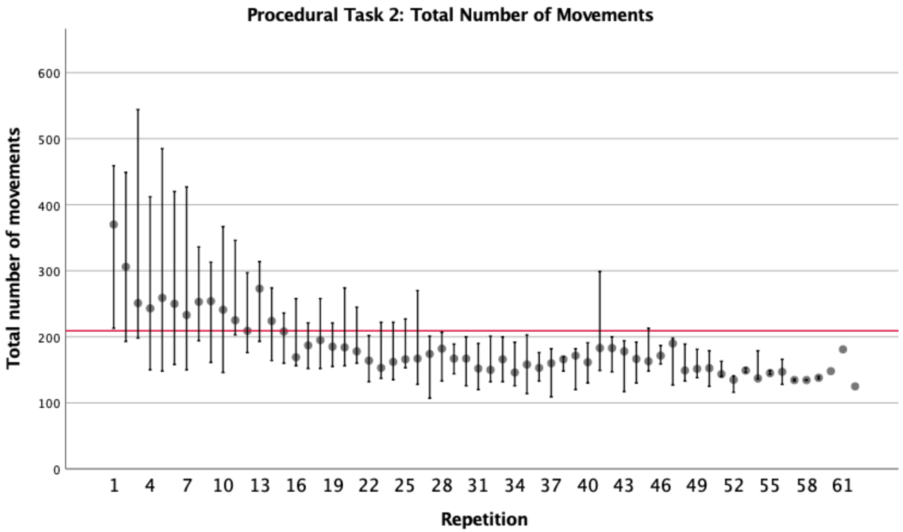

Psychomotor Task 9: Total Number of Movements
Total number of movements
0
500
1,000
1,500
2,000
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70
Repetition

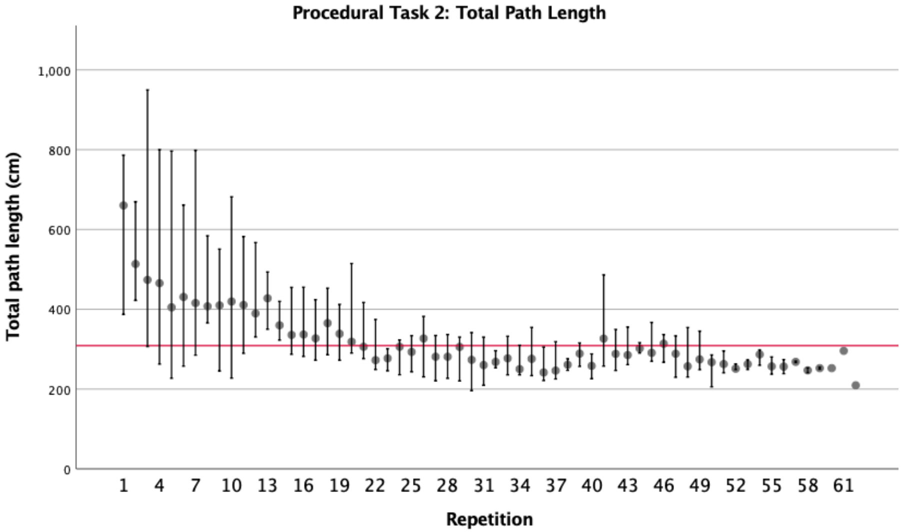

Psychomotor Task 9: Total Time
Total time (seconds)
1,000
800
600
400
200
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70
Repetition

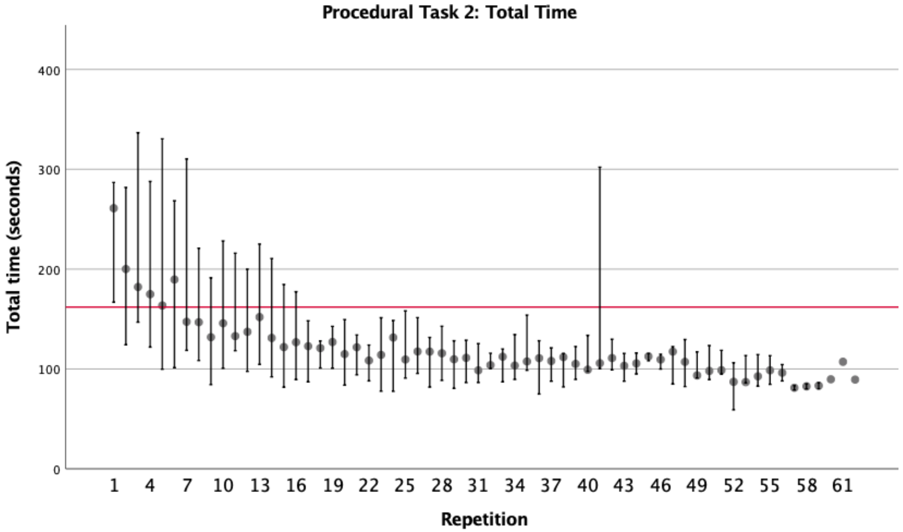

Procedural Task 1: Total Number of Movements
Total number of movements
800
600
400
200
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97 100 103 106 109 112 115 118 121 124 127 130 133 136 139 142 145 148 151
Repetition

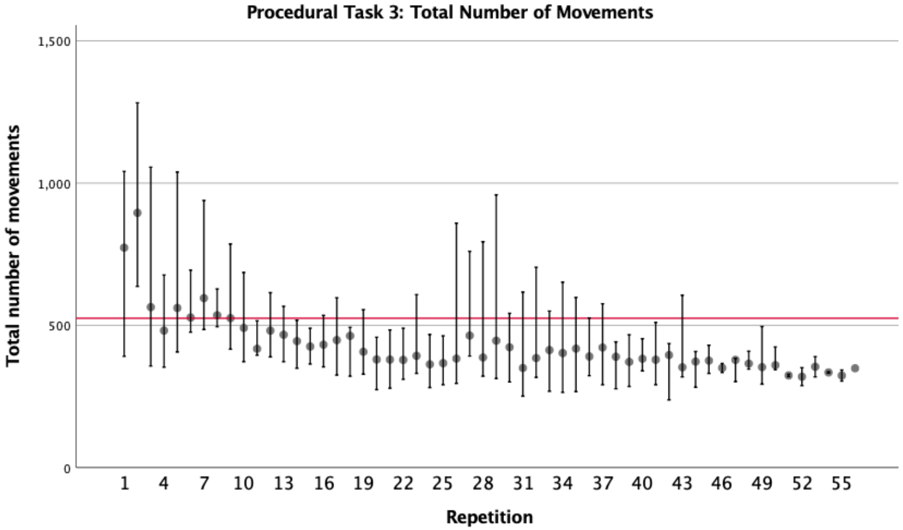

Procedural Task 1: Total Path Length
Total path length (cm)
1,200
1,000
800
600
400
200
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97 100 103 106 109 112 115 118 121 124 127 130 133 136 139 142 145 148 151
Repetition

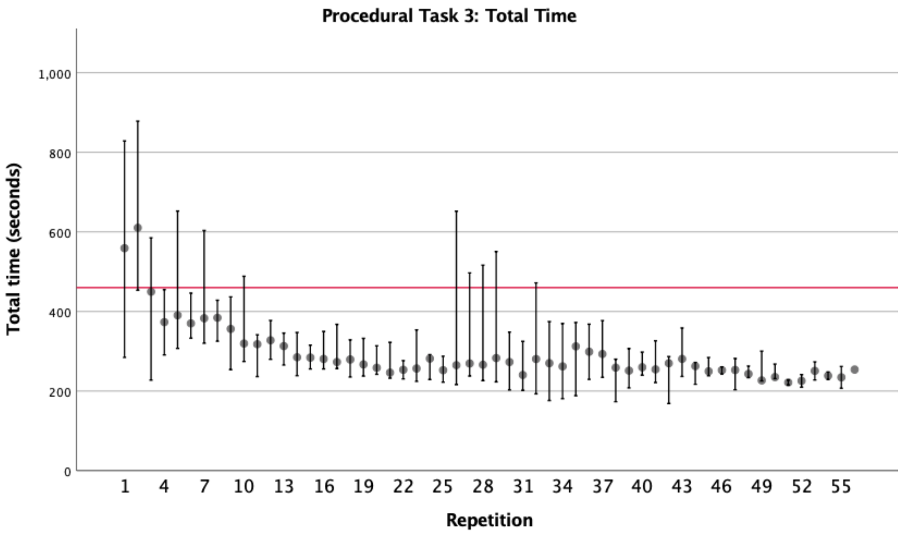

Procedural Task 1: Total Time
Total time (seconds)
400
300
200
100
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82 85 88 91 94 97 100 103 106 109 112 115 118 121 124 127 130 133 136 139 142 145 148 151
Repetition


Procedural Task 2: Total Number of Movements

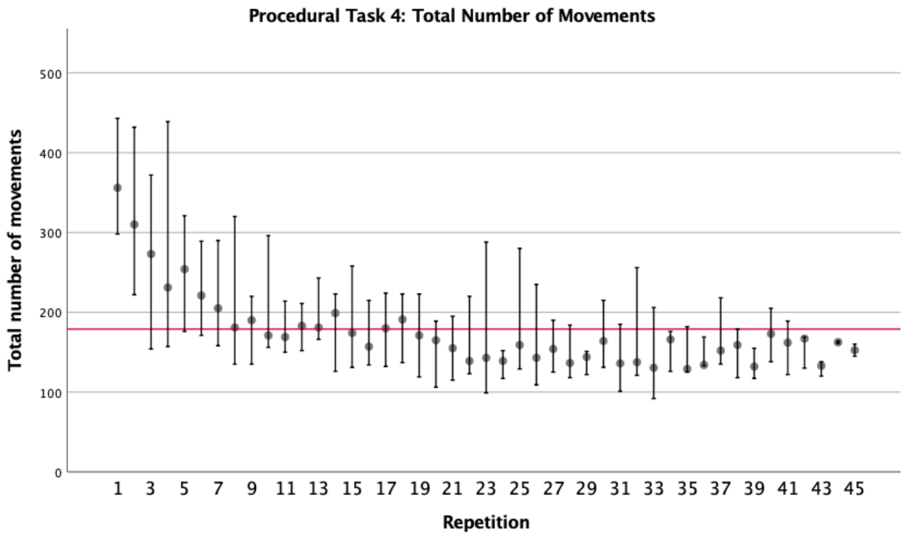

Procedural Task 2: Total Path Length


Procedural Task 2: Total Time

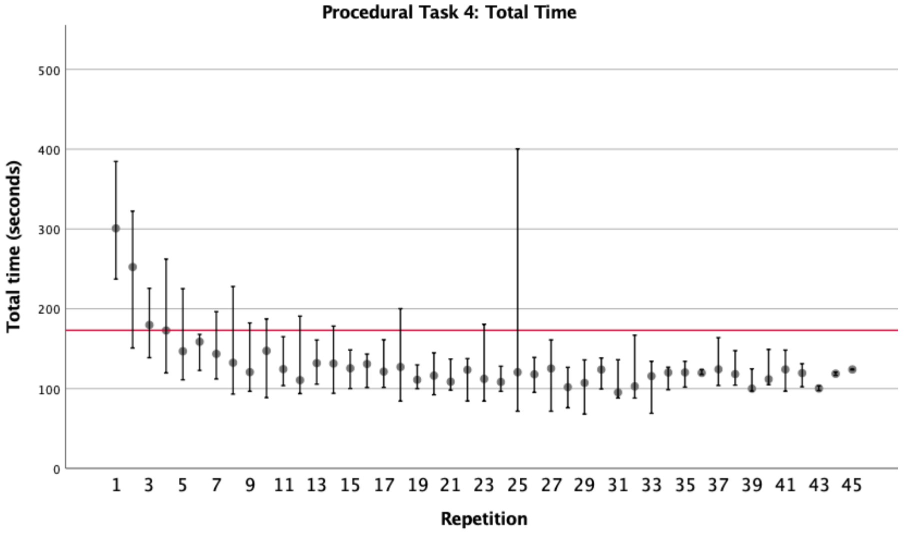

Procedural Task 3: Total Number of Movements



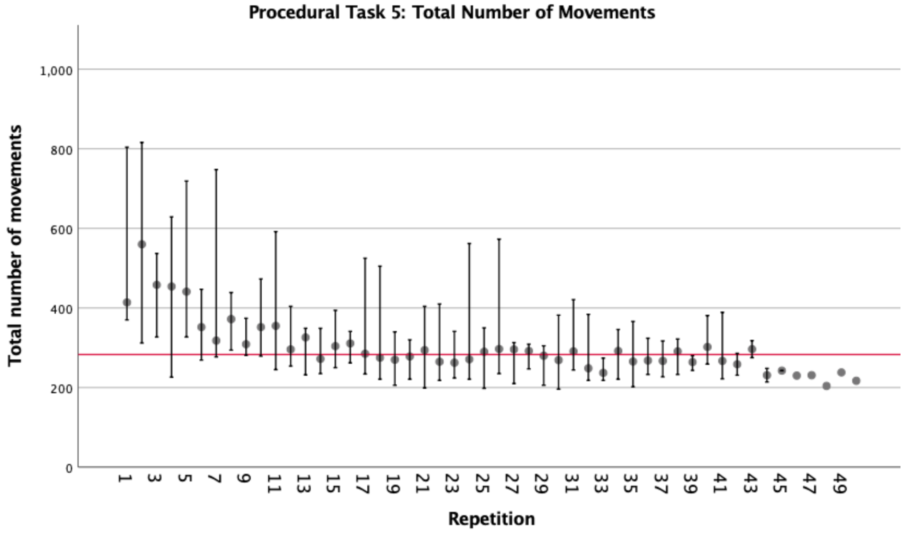
Procedural Task 3: Total Time



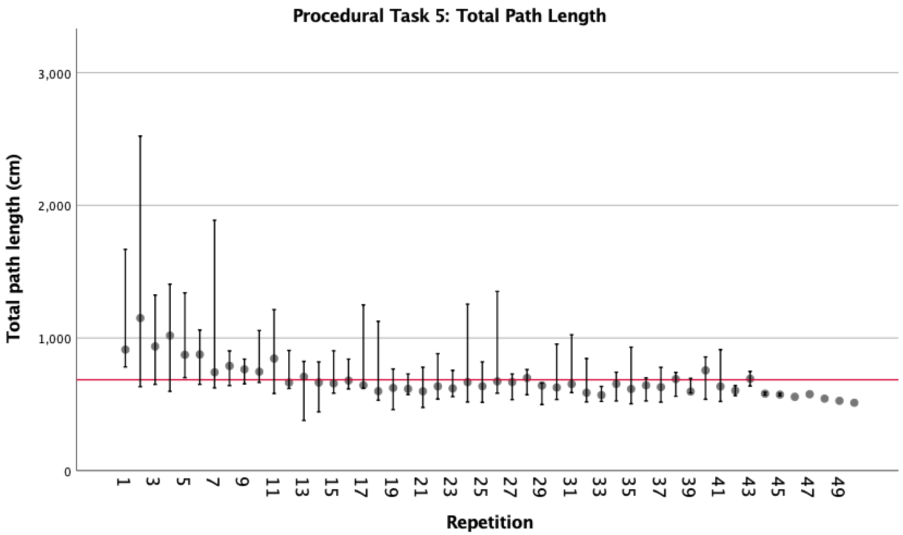
Procedural Task 4: Total Number of Movements

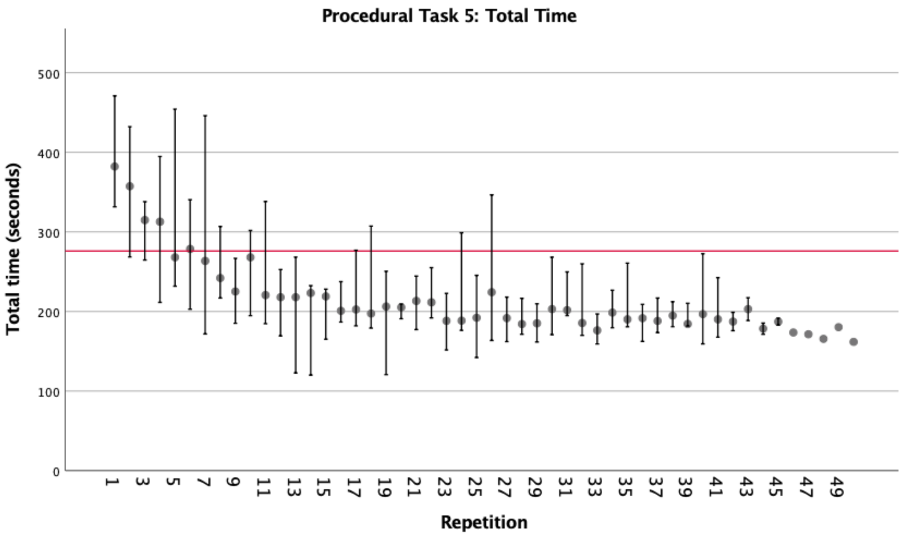


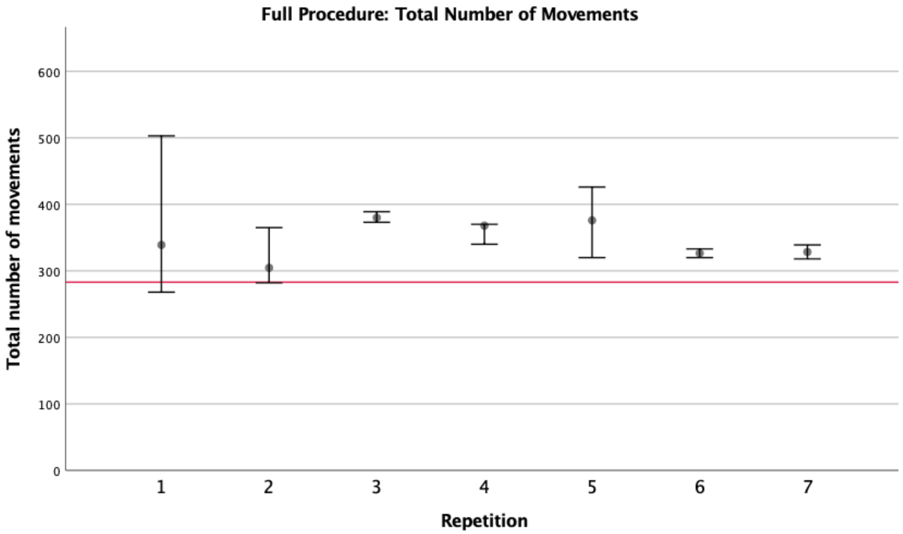
Procedural Task 4: Total Time

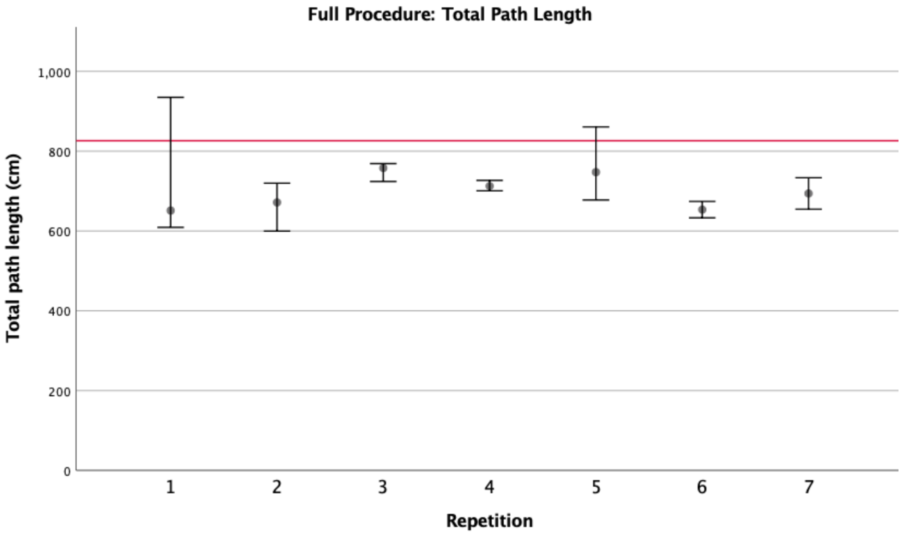


Procedural Task 5: Total Number of Movements


Procedural Task 5: Total Path Length
Total path length (cm)
0
1,000
2,000
3,000
Repetition
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49


Procedural Task 5: Total Time
Total time (seconds)
0
100
200
300
400
500
Repetition
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49


Full Procedure: Total Number of Movements
Total number of movements
0
100
200
300
400
500
600
Repetition
1
2
3
4
5
6
7


Full Procedure: Total Path Length
Total path length (cm)
0
200
400
600
800
1,000
Repetition
1
2
3
4
5
6
7

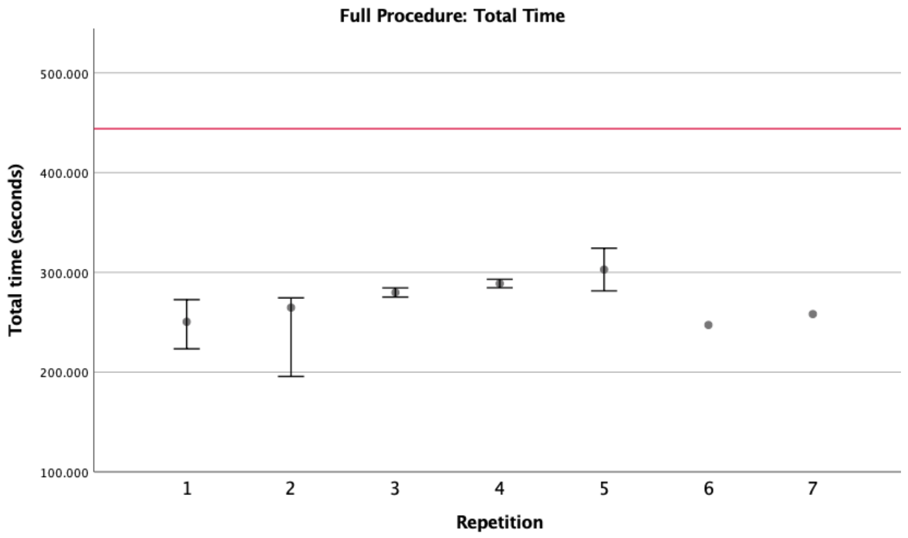


Figure 11 - Learning curves for all metrics of the Laparoscopic Appendicectomy curriculum (excluding error scores)*. Median with 95% confidence intervals are displayed. The red line corresponds to the performance level required for progression as specified in the curriculum.*

Participant	Psychomotor Task 5	Psychomotor Task 9	Procedural Task 1	Procedural Task 2	Procedural Task 3	Procedural Task 4	Procedural Task 5	Full Procedure
1	50	57	153	59	35	31	27	2
2	21	22	21	37	25	27	24	2
3	40	38	74	37	31	33	35	5
4	91	70	66	51	48	45	45	7
5	59	69	72	62	54	43	41	1
6	27	55	112	55	50	45	50	7
7	65	58	50	35	22	29	22	2

Table 19 – Number of repetitions of each task of the Laparoscopic Appendicectomy curriculum, by participant.

	1	2	3	4	5	6	7
Depth perception	5	4	5	5	5	4	4
Bimanual dexterity	4	5	4	5	5	5	4
Efficiency	4	5	4	5	5	5	5
Tissue handling	4	5	3	3	5	5	5
Autonomy	5	5	5	5	5	5	5
Habits	4	5	5	5	4	4	5

Table 20 - Final assessment scores using the modified GOALS tool. *1 = poorest, 5 = best..*

Participant	Feedback
1	Efficiency: 4 rather than 5 as had a short period of time repeatedly dissecting the mesoappendix at the distal aspect of the appendicular artery. Tissue handling: 4 rather than 5 as inadvertently 'brushed' the terminal ileum on initial entry into the abdomen. Habits: On one occasion entered and withdrew the scissors with the tips kept open. Overall, demonstrated progressive improvement at each session, in particular reducing number of movements to complete dissection of the mesoappendix and accurate placement of endoloops. In the future, aim to improve use of left hand to alter tension or change angle of approach.
2	Dissection of appendicular artery: Strand of mesoappendix behind the artery left after clipping & division of artery. Dealt with promptly and safely by application of bipolar diathermy. Dissection of appendix base: Could have cleared a greater area distally to allow a larger distance between the proximal and distal endoloops as the second proximal endoloop would have fallen off in real life after division of the appendix. Excellent use of roticulators Good tissue handling, consistent application of appropriate force. Slick and accurate placement of endoloops. Overall, a smooth and safe operation, well done.
3	Efficient use of reticulation. Path length very occasionally "jerky" rather than a single smooth movement. Careful dissection of window – good technique. Somewhat rough at points in the dissection of the edges of the artery – placed it under undue tension – risk of rupture in real life. Indented (i.e. pushed instrument into) terminal ileum with scissors upon one entry into the abdomen. Very accurate placement of endoloops. Overall a safe and efficient operation.
4	Smooth use of roticulators – conjugate movements. Focused, atraumatic dissection of the mesoappendix. Excellent use of left hand to alter the view when needed. ?Incomplete ligation of artery: minor bleeding from distal artery. Accurate placement of endoloops – appropriate spacing proximally, and divided the appendix to leave a "cuff" to hold the endoloops. Overall an efficient and smooth operation. Very well-controlled path length no unnecessary movements. Unfortunately, ?did not appear to completely ligate the appendicular artery, but this was quickly controlled (and also did not become unduly flustered) – could well have been an artefact.
5	Safe, efficient dissection of mesenteric window and artery. Smooth movements, good use of roticulators. One ?entry with scissors in left hand with tips open. Appropriately spaced endoloops and size of division of the appendix. Left hand used to set tension correctly throughout or alter view if necessary.
6	Overall a slick, safe and efficient procedure. Targeted & accurate creation of mesoappendix window. Excellent use of the left hand to adjust the view. Occasional entry into the operative field with [blunt] instrument tips open. Intermittent minor inaccuracies in depth perception when dissecting at the appendix base, also at this point could have slowed down a little to reduce repetitive movements. "Textbook" application of endoloops – a very careful and progressive reduction in the loop size.
7	Dissection: Gentle and careful, however when dissecting out the artery from the mesoappendix could have applied greater force to more efficiently isolate the vessel (i.e. fewer movements). Use of left hand: Generally good, kept both instrument tips in view at all times, however could have altered the view more often using the left hand to aid visualisation e.g. when clipping and dividing the artery. Basse of appendix: Could have removed more fat (of the mesoappendix) right at the appendix base to ensure secure ligation of the appendix Excellent endoloop application. Overall, a safe and deliberate operator. Maintains task focus, efficient and at ease with the instruments used. Well done!

Table 21 - Individual feedback recorded by the examiner as part of the final assessment.

3.3.1 Participant Diaries – Thematic Analysis

All seven participants completed a diary over the course of their participation and these were subject to thematic analysis. This converged upon three distinct themes: frustration, feedback sessions, and simulator access and attendance.

3.3.1.1 Frustration

Participants expressed frustration within two sub-themes: curriculum progression and simulator failures.

3.3.1.1a Curriculum Progression

Participants often expressed frustration either as a response to acute events or related generally to their progression along the curriculum. For example, frustration was often felt acutely due to occasional mistakes and momentary errors such as dropping objects and clips. An underlying and more general sense of frustration was borne from the overall perception of how difficult it was going to be to complete the curriculum. It was frustrating for participants to realise how many repetitions were being required and to experience repeatedly missing the progression criteria. Psychomotor Task 9 was frequently identified as being particularly difficult. Disheartenment and frustration grew upon comparing gaps in performance with the progression criteria, which often led to feelings of 'giving up' and feeling that the criteria may be unachievable.

However, frustration was frequently countered by feelings of great rewards, satisfaction and excitement upon meeting the fulfilment criteria for each section. Participants were reassured by knowledge of others' similar experience and one participant was gratified and encouraged by his performance in the real OR. It was also notable how the feelings of frustration were countered by the realisation of a *sunk cost*, which led to feelings of needing to persevere.

3.3.1.1b Simulator Failures

Software errors were frequent enough to cause notable frustration and disruption amongst participants. Errors either caused the entire simulator to crash during a repetition, necessitating a restart of the entire machine, or they caused recurring glitches. Such glitches included:

- the Endoloop passing through the appendix as it was being tightened;
- firing of the stapler not being recognised by the simulator;
- a missed clip being registered if attempting to clip a tube that has no water flowing from it (Psychomotor Task 5);
- difficulties encountered by the simulator in registering shapes that had been placed on the ground appropriately (Psychomotor Task 9);
- instruments being stuck within objects;
- not being able to place the appendix in the Endobag;
- unpredictability in the magnitude of dissection sometimes required for the simulator to move on from the green-marked area of the mesoappendix.

> *‘Due to a glitch the loop caught on the part of the colon, on another occasion it has gone through the appendix when I attempted closing the loop. I am finding this extremely frustrating.’*

The *green glitch* was a considerable source of frustration. It led to participants feeling that they were trying to please the machine and find ways to ‘cheat’ it in order for the simulator to allow them to progress.

> *‘Almost feels as if one is trying to jump through hoops rather than completing a task to satisfaction like real surgery.’*

3.3.1.2 Feedback Sessions

The feedback sessions were unanimously held in high regard – they were felt to be very helpful with participants ‘needing’ and ‘looking forward’ to them, and they felt encouraged by the coaches. Whilst these sessions were a helpful reminder of the steps for each task, there were deeper benefits experienced by participants. For example, frustration was countered as coaches were able to demonstrate that the completion criteria were achievable. Indeed, knowing that feedback sessions were available seemed to provide reassurance that there was another source of performance improvement when patience was running low and frustration was rising.

Observing the coaches enabled participants to identify differences between what the coach was demonstrating and their own technique. Furthermore, participants consistently provided examples of how the coach pointed out areas of weakness that had previously gone unidentified by the participant during self-directed practice. Participants were equipped with specific knowledge of how to improve, including unsafe habits related to use of the instruments that the metrics will not have reported, and they were encouraged to concentrate on tasks where performance was weaker. In this regard, coaches played an important role in encouraging more deliberate practice by the participants.

> *'Addressed my weak points. And I noticed a lot of things I was doing incorrectly, especially with the location of my dissecting. Instructor showed me places to improve and it feels positive that there is actually a possibility of improving - as previously it was looking impossible.'*

Instructor sessions appeared invaluable with respect to confidence and motivation of the participants. Positivity and enthusiasm were often renewed for tasks that were felt to be impossible to complete successfully, by providing a sense that further progress was possible. One participant was greatly motivated having been shown a video of how a previous participant had successfully completed a task.

3.3.1.3 Simulator Access and Attendance

Open access of the simulators allowed participants to fit practice sessions around their individual schedules. They were motivated to attend whenever they could, feeling that it was 'lots of fun', with 'actual surgical simulation which makes it more exciting' and one participant stating that she or he 'really looked forward to doing the session'. Excitement and motivation had also been directly linked to progression onto the LA tasks, as it resembled real surgery more closely than the BS tasks. There were suggestions that such open access occasionally led to clashes when machines were in use, with delays waiting for a simulator to become available having the potential to feel distracting.

There were feelings of fatigue when attending after night shifts, which may have affected concentration, with a sense that more mistakes were being made on the simulator as a result. Participants were often juggling sessions between clinical duties and one participant reported having to leave halfway through a session due to being on call. Participants indicated that 45 minutes was the correct length of time for each session, influenced by fatigue, concentration span and accumulating frustration.

> *'I felt tired and was quite relieved to be leaving at the end of the 45 minutes.'*

3.3.2 Data Saturation

Analysis of the fifth and sixth participant diaries yielded no new themes, and so the New Information Threshold of ≤5% was reached.

4 Discussion

This study is a prospective randomised assessment of the LAP Mentor appendicectomy module, with all tasks demonstrating construct validity together with several of the BS tasks. Performance curves were also significant for the majority of construct-valid metrics, which is important for demonstrating that performance can improve towards the proficiency levels set by experienced participants by practicing on the simulator. This informed the construction and implementation of a structured VR LA curriculum. Both novice and experienced participants perceived the LA module to be realistic and useful for developing laparoscopic skill.

This study also demonstrated the first assessment of real-world implementation of a VR laparoscopy curriculum that also included a qualitative analysis. This revealed some dropout of participants and perceptions of a difficult curriculum that required many repeat sessions in order to complete successfully. The suggestion that the level of difficulty was responsible for the dropouts is circumstantial, and unfortunately the reasons remain unknown. There was notably high utilisation and regard for the feedback sessions.

4.1 Evaluation of the Assessment of Construct & Face Validity

Our findings of construct validity in the BS tasks partially corroborate those of others (96,107–109). However, there are some notable exceptions, such as Aggarwal *et al*'s construct validity results for BS task 5 (96) being limited to *total time* and *total speed*, whereas in the present study the majority of BS task 5 metrics demonstrated construct validity. Matsuda *et al*'s study (108) correlated accuracy in BS tasks 4, 5 and 8 with blinded expert assessments of real laparoscopic procedures (construct validity was not demonstrated for accuracy in BS task 3), though in the present study only accuracy in BS task 5 demonstrated construct validity. Other notable variations include Aggarwal *et al*'s construct validity demonstration for BS tasks 1 and

2 (96), unlike the present study, or that by McDougall *et al* (109), which used composite performance scores and demonstrated construct validity for BS tasks 3 to 9 only. Wilson *et al* (99) studied BS task 3 and found construct validity for *total time* and *economy of motion (left)* and Yamaguchi *et al* (100) also demonstrated construct validity for these metrics in addition to *number of movement (left)* and *average speed (left)*, whereas the present study instead demonstrated construct validity for *relevant path length (right)*, *total path length (right)* and *economy of movement (right)*, and not for *total time*. Such inconsistencies in the literature may be related to significant variations in laparoscopic experience that are evident between novice groups, and in the repetitions used for statistical analysis (110,111).

The assessment of construct validity in the present study required novices to complete 10 repetitions of their assigned task(s), which is greater than the number by which novices have been able to reach expert-level proficiency in previous studies of VR laparoscopy (96,101–103). However, the shape of some of the learning curves in the first part of the study suggest that ongoing practice may have resulted in further improvements in novices' performance. This has been corroborated by others, with performance improvements and multiple plateaus seen towards the 30th repetition (112). During the Part 1 of the present study, some appendicectomy tasks in our curriculum feature metrics where expert performance was not actually reached by the 10th repetition. However, Part 3 reassuringly demonstrated that expert proficiency was still achievable.

Unfortunately, system errors resulted in the loss of many data points for BS tasks 4 and 6 in the present study, which may have resulted in underpowered analysis of some movement metrics, although it is notable that all but one metric (*ideal path length, left*) did not demonstrate construct validity. Besides this, our study protocol was strengthened by a unique fully-supervised training regimen during this first part of the study that ensured adherence to a strict protocol where instruction and feedback were prohibited, guaranteeing the integrity of our data. Another unique strength is the randomized, prospective study design that aimed to reduce the warm-up effect on later tasks in novices and experienced participants by separating them *both* into task- or module-specific groups.

4.2 Design of the VR Appendicectomy Curriculum

To date, construct validity has been demonstrated in procedural modules across a variety of VR simulators, such as the MIST nephrectomy (97), LapSim salpingectomy (103), LAP Mentor salpingectomy/salpingotomy (98) and the LAP Mentor cholecystectomy modules (96). The present study is the first to develop a curriculum for the LAP Mentor appendicectomy module that is based on a strong theoretical framework, incorporating deliberate practice and formative feedback. This curriculum encourages learners to reflect and consciously seek to improve specific aspects of performance with each repetition, to achieve expert performance rather than arrested development (30). It seeks to achieve this by a) including tasks that have been deliberately selected to improve skill (44), b) providing specific goals, which accelerates learning during VR training (11,23), c) providing immediate feedback (performance metrics), and d) being self-directed and tailored, allowing trainees to rehearse in their own time and directing them to repeat and concentrate on tasks that are not yet expertly performed.

In the VR appendicectomy curriculum of the present study, we included core psychomotor skills tasks in order to create a gradual, step-wise increase in task complexity, which was deemed an appropriate strategy for a curriculum that is aimed at surgical novices. Tasks and metrics that demonstrated construct validity were included with scores from experienced laparoscopic surgeons as goals. Error scores were included despite not demonstrating construct validity, since it is important that speed is not rewarded at the expense of error.

Another notable feature of our VR appendicectomy curriculum is its design for distributed rather than massed practice (favouring spacing intervals between practice sessions), as it is now well-documented that this results in superior learning and retention of skill (68). As outlined earlier, spacing intervals must be considered with respect to the importance of sleep and its possible role in the consolidation of learning (the *memory consolidation hypothesis*) and the potential detriment of continued practice during this process (66,67). The optimal spacing interval is not yet known and may depend on task type (68), though so far there is evidence to suggest that daily or weekly sessions may be the most efficient for learning and retaining psychomotor skill (69).

A unique feature of our curriculum is the mandatory periods of human feedback, and it should be acknowledged that the availability of faculty is a potential limiting factor that fortunately did not affect implementation of the curriculum during the study period. The timing of the delivery of human feedback in the curriculum falls in line with our current understanding of the relevant science, by providing it intermittently (summary feedback) rather than continuously during each repetition (concurrent feedback), as has been discussed earlier. Participant diaries have demonstrated that coaches had directed participants to weaker aspects of performance, which is a core tenet of deliberate practice.

Qualitative data from participants diaries have also revealed the important role coaches may have in countering frustration that is borne by attempting a task that is just beyond the zone of optimal motivation – if a task is too easy it may elicit boredom, whereas a task that is too difficult may elicit stress and frustration (115). Coaches were crucial in motivating participants and countering their frustration during what is evidently a very challenging curriculum.

Uncertainties exist surrounding the optimal session duration, the maximum number of sessions per day and the criteria for allowing participants to progress through the curriculum. So far, only expert opinion without high-quality evidence has suggested that there should be no more than two training sessions per day (116,117) with each being limited to one hour (117,118) in order to limit mental fatigue (119). Many subscribe to this training regimen in their curricula, together with an accepted arbitrary criterion of having to demonstrate proficiency in two consecutive attempts before trainees are allowed to progress to the next stage (56,96). Clearly, these curriculum design features require further study. The present investigation provides qualitative evidence from participant diaries that 45 minutes is likely to be the upper limit for the optimal session duration, though it may be influenced by other design features such as task type, level of difficulty, and the time of day of usual participation.

Others have developed similar VR curricula for other procedures, such as laparoscopic cholecystectomy and salpingectomy, which have crucially demonstrated translation of VR performance to the real operating environment (25,120). In one randomised and controlled transferability study, Ahlberg *et al* demonstrated significant and consistent improvements in the performance of 10 consecutive laparoscopic cholecystectomies following VR training (120). In another landmark randomised trial, Larsen *et al* demonstrated better and quicker real salpingectomies in the VR procedural-trained group versus control (25).

4.3 Implementation of the Curriculum – A Closer Look

Implementation of a curriculum is critical to understand real-world aspects such as its actual utilisation (including practical issues such as simulator access and the availability of faculty), achievability of the objectives, and participant motivation. Such assessment was undertaken by combining mainly qualitative with some quantitative methodologies.

Performance curves of the participants were remarkably long, which was not anticipated. Unfortunately, it is not known whether the three dropouts were linked to this. The learning curves suggest that most of the performance gains occur within the first few repetitions. A long tail to these curves is probably a result of having to fulfil more than one performance criterion within the same repetition, and then needing to repeat this in consecutive attempts across more than one task in order to progress. This requires, by definition, reduced performance variability (which is highly desirable), and there may be some suggestion of this by the narrowing confidence intervals that can be seen as the learning curves progress. However, such quantitative analysis fell outside the study’s primary objectives and so it was not powered nor designed to allow for any meaningful statistical analysis in this regard.

For qualitative analysis, however, the high level of thematic saturation seen in participant diaries suggests that a larger sample is unlikely to have yielded any new information regarding participants’ experiences of the curriculum. Indeed, the use of participant diaries, seldom seen in the literature, yielded critical insights regarding its implementation.

The qualitative feedback in these diaries suggested that participants were excited at various points along the curriculum, though they did feel that it was difficult and clearly required a lot of practice. However, it is likely that this contributed to a greater sense of satisfaction once it was finally completed – a sense of reward was frequently documented.

Frustration was a significant component of participants' experiences, and the time and effort required to complete the course mean that motivation is critical. Frustration resulted from software glitches affecting fidelity of the simulators, which were frequent. Complete crashes also occurred in the middle of repetitions, necessitating a complete reboot of the machine at critical moments. Perhaps more concerning were frequent reports of participants feeling like they had to simply please the 'machine' rather than technically perform as best as possible. This was particularly evident in PT1 (seemingly the most difficult task by measure of the number of times it was attempted), where participants tried to learn a specific order of dissection within the green-marked area that would result in the computer registering it as complete. Clearly these are technical issues that ought to be addressed by the manufacturer.

Frustration also arose from a sense that the tasks were simply too difficult and that the objectives were unachievable. This inevitably affected participants' motivation. However, this is where feedback sessions proved invaluable.

Firstly, coaches facilitated deliberate practice – immediately, participants were encouraged to concentrate on the tasks that were most difficult (following the prescribed order only when attempting to meet the criteria that would allow them to progress to the next stage). They were also frequently able to identify weak areas of performance and demonstrate new ways to find performance improvements to which participants were initially blind. Not only is this an important component of deliberate practice towards mastery, this aspect was central in the mitigation against frustration and losses of motivation.

Secondly, there is clear evidence from participants' diaries coaches were teaching them safe habits that the simulators do not address, such as avoiding unnecessary contact with bowel, and the safe entry of instruments into the abdomen. This is also evidenced by the high scores in the *habits* domain of the modified GOALS tool, though it must be acknowledged that this aspect of the assessment has not been validated.

Participants' diaries also support the open access policy of the simulation centre – being able to work around clinical duties and work schedules appears to have greatly facilitated participation, although this ‘juggling’ allowed fatigue to creep into their experience. However, this is arguably acceptable since it better matches real-world conditions, where surgeons are required to perform at various levels of alertness, and not just when well-rested.

Finally, it is worth considering the impact of the participant diaries themselves on performance. The ‘observer effect’ in particle physics relates to how observation itself can affect the observed reality (121). Analogously, given the potential benefits of written reflections upon learning and development (122), it is plausible that participants’ writing in diaries may have positively affected their performance. Given that all obliged in this reflective activity, it may be argued that this aspect should be continued beyond the present study – to facilitate both learning and the continual development of the curriculum through its value as a real-time feedback tool.

4.4 Existing Validity Evidence & Curricula for VR LA

The findings in the present study in support of training to perform LA in the VR environment have been corroborated by others. Some have also developed their own curricula, though with notable differences to that of the present investigation, but no research has been conducted to evaluate their implementation.

Sirimanna & Gladman split 38 participants into three groups according to their level of experience, who then underwent ‘baseline skill testing’ (91). The median performance during the second repetition was used to assess construct validity for each of the LAP Mentor procedural and FP tasks. This contrasts the present study, which used the first two repetitions and no baseline skill testing. Intermediate and experienced participants also completed all baseline skill testing, procedural and FP tasks sequentially, all contributing to the possibility of significant warmup effect, whereas the present study’s strength lies in the randomisation and separation of both experts and novices between procedural and FP tasks.

In spite of this, the results were broadly similar, with total path length, total movements, total time and idle time demonstrating construct validity across all appendicectomy tasks, together with demonstrable learning curves. A curriculum was designed that included all nine BS tasks rather than selecting only those that have specific content and construct validity, and with no instruction or coaching (content validity suggests relevance of the task to the performance of LA). Unfortunately, none of the data that were collected to assess face and content validity was presented. Further differences to the present study include the inclusion criteria for novices, which are generally heterogenous amongst such studies. In comparison, the present study included much more experienced participants in the 'expert' group, and a novice group that had no experience at all of performing laparoscopic surgery. Our selection criteria for novices were designed to be as representative as possible of those for whom the curriculum has been designed.

More recently, Nayar *et al* (2019) assessed aspects of laparoscopic surgery that are seldom studied. The NASA Task Load Index (NASA-TLX) and Subjective Mental Effort Questionnaire (SMEQ) detected significant differences in mental loading between experienced and novice laparoscopic surgeons. In addition to demonstrating construct validity in time and movement metrics, the second session of each participant was scored by blinded experts using the Objective Structured Assessment of Technical Skill (OSATS) tool. This contributed to the validity evidence of the LAP Mentor LA module by demonstrating construct validity not just in the quantitative dexterity parameters, but also qualitatively and in the cognitive aspects of performing complex psychomotor tasks (123). The content validity of the module was also supported since there was strong agreement by experienced participants with a series of statements that were derived from a modified version of a previously published Cognitive Task Analysis (CTA) script that was specific to LA (123,124).

The most interesting part of the results was the reduction in mental loading at the seventh session, which strongly correlated with total time, total number of movements and idle time (Spearman's rank correlation coefficient = 0.90). The suggestion was that the automaticity that is developed by training on a simulator would allow the trainee surgeon to direct his or her attention to other aspects of the procedure, such as the decision-making (123).

The LapSim (Surgical Science, Göteborg, Sweden) is another high-fidelity simulator with haptic feedback and its own appendicectomy module. Bjerrum *et al* (2017) split 45 participants into three groups by their level of experience, all of whom performed 20 repetitions of a full unguided appendicectomy following a familiarisation session. Significant learning curves were demonstrated and construct-valid metrics included right and left instrument tip path length, and total task time (125). Number of movements is not a metric that is provided on the LapSim module, which is unfortunate given how it has been shown to consistently differentiate between novices and experienced surgeons on the LAP Mentor.

Brown *et al* performed an experiment whose real value is in demonstrating the problems that can arise when investigating simulators for validity. The 'novice' group of surgeons had already performed a mean of 13 laparoscopic appendicectomies (range 0 – 55), and data collection was performed at a 'hybrid boot camp blended with clinical operative training', both aspects that introduce significant bias and limit the applicability of the results. Training grade was significantly associated with a better composite score, for which learning curves were constructed (126). The problem with composite scores is that it is very difficult for the trainee to understand how to improve their performance when presented with such a meaningless number, making practice less 'deliberate' than it ought to be.

Whilst demonstrating construct validity for this composite performance score as well as total time, Beyer-Berjot *et al* integrated the LapSim appendicectomy module into a novel perioperative virtual care pathway. Participants were required to assess and manage specially-designed cases in a virtual world, following them through to the virtual OR where they performed their surgery on the LapSim, and then manage these virtual patients post-operatively. For the operative aspect, a curriculum was constructed that was based on a series of psychomotor tasks followed by the LA tasks (127). Though the main aspect of the study appeared to be around the design of the virtual care pathway, it only examined the operative aspect for validity. It is an interesting attempt at providing a holistic and integrated approach to laparoscopic surgical training, though it may be better suited for more complex cases that present much less frequently compared to appendicitis. This is simply because it is very easy for a trainee to find real patients with right iliac fossa pain and become involved in their pre- and post-operative aspects of their care in a real environment that has, by definition, complete content validity. Indeed, there may be value in incorporating simulation curricula into the wider aspects of appendicectomy in a real hospital environment. For example, in

Norway's Stavanger University Hospital, where mandatory simulator training and didactic teaching on the peri-operative aspects of appendicectomy was followed by real laparoscopic appendicectomies by trainees who were given structured feedback after each procedure nearly two thirds of the time (128).

In summary, the results of the present study are supported by others, albeit with differences in methodology that become important when considering the group at whom the module is targeted. Some have also made attempts at constructing VR curricula for LA. However, the present curriculum is more focused on the LA module itself rather than the BS tasks, which would lower the risk of construct-irrelevant variance (including exercises that are not relevant). This curriculum appears to be more challenging since the number of repetitions required before proficiency is markedly greater than that seen in these aforementioned studies. Ultimately, it is difficult to make direct comparisons between curricula. To our knowledge, the present study is the only one to have implemented a curriculum for real-world training, and with data that contributes to the validity evidence of the curriculum as a whole rather than just the simulator modules themselves. The strengths in the present curriculum are 1) that it is supported by a robust theoretical framework and 2) that it includes regular formative feedback sessions as a cornerstone of training.

4.5 Theoretical Framework Controversies

The theoretical frameworks underpinning the validation of simulators and skills curricula have been previously discussed. The older APA terminology has been utilised in the present study for reasons already alluded to, for example: construct validity was used because the context of the study was *training* rather than *testing*, the construct is believed to be relevant and expert performance metrics were also believed to be eventually attainable by novice trainees. Furthermore, there is consensus: the majority of the published literature utilises the same APA framework and terminology (88), including SAGES and the widely-recognised FLS curriculum (92). Finally, the chosen framework is arguably the easiest to understand and communicate (90).

Table 22 demonstrates how the findings of the present study are transferable and can still be considered as sources of evidence according to the Messick Contemporary Framework. Unlike construct validity, whose place in it is relatively easily understood, the role of face validity as a source of evidence is much less clear. Borgerson *et al* believe that it is not transferrable to the Messick Contemporary Framework (88). Although face validity does not appear in the APA framework (its origins are unclear), it is still considered an acceptable source of validity according to 41.6% of papers reviewed by Borgerson's group (88). Without face validity there could be a greater risk of construct-irrelevant variance. It is also relevant since it is difficult to imagine how a curriculum would garner confidence from trainers, investment of time by trainees and financial investment from institutions, without the simulator bearing any resemblance to reality for which it purports to reflect and train. In fact, the qualitative analysis performed in the present study uncovered direct links between the fidelity of the simulator exercises to real surgery (in other words, its face validity) and participants' excitement and motivation.

Source of Validity	How is this addressed?	How could it be addressed in future study?
Content	14 of 17 experts rated their corresponding part of the appendicectomy module as 'useful' for developing the skills for a 'safe performance of real laparoscopic appendicectomy' in the 'post-course' questionnaire. Experts also believe that a course based on this module should be mandatory[†].	
Response Process	Data collection (from which curriculum items were set and proficiency scores were calculated) was highly standardised and participants were supervised throughout. Curriculum's proficiency criteria (apart from final assessment) are objective. Curriculum's final summative assessment by modified GOALS.	
Internal Structure		The curriculum's final assessment of trainees' performance could involve multiple experts in order to determine inter-rater reliability.
Relations to other variables	Significant differences in performance were found between novices and experts (theoretical relationship to test scores: the level of experience performing real laparoscopic appendicectomy).	
Consequences of the test		Future study to determine whether those who have completed the curriculum perform better in real laparoscopic appendicectomy, or in real laparoscopy in general. For example, blinded assessment of real trainee performance.

Table 22 – Sources of validity of the Laparoscopic Appendicectomy module and curriculum, according to the Messick Framework. *The evidence gathered from the present study has been categorised according to this framework (88).*

4.6 Frustration in Surgery

Frustration was a common theme reported by participants during completion of the curriculum. This was a particular issue during the second stage, where participants are required to complete each of the five procedural tasks within the set criteria, twice in a row. Despite the frustration, all participants who experienced this were sufficiently motivated to continue the course and successfully complete it. This is encouraging, but it highlights the non-technical aspect of stress and frustration in surgery, which was not measured during the study period.

It is important to distinguish 'frustration' from 'stress' in order to better understand their triggers and effects on surgical performance. Frustration can be defined as

> *'a key negative emotion that roots in disappointment (Latin frustrā or "in vain") and can be defined as irritable distress after a wish collided with an unyielding reality'* (129)

whereas stress can be described as

> *'a particular relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and endangering his or her well-being.'* (130)

In the context of surgery, frustration is experienced, for example, when the 'wish' of grasping a thread accurately in order to tie a knot laparoscopically collides with the 'unyielding reality' of repeatedly failing to orientate the thread, resulting in a very challenging experience in the subsequent tying of the knot. Or in the case of completing our curriculum, participants experienced frustration when they were repeatedly faced with the reality of having to re-attempt a series of tasks due to the reality of not meeting the performance criteria.

Frustration in surgery has seldom been studied. In 2005, Kaafarani *et al* randomised 1,983 patients to receive either open or laparoscopic hernia repair (131). After the procedure, surgeons were asked to rate their frustration during the procedure, and their overall satisfaction with the result. Interestingly, surgeon satisfaction did not correlate with outcomes, which corroborates the results of other similar studies that were albeit limited to orthopaedic procedures (132). However, frustration during hernia repair was associated with poorer outcomes. Hernia repairs where the surgeon reported any frustration were twice as likely to recur within two years compared to procedures where no frustration was reported. These patients were also 30% more likely to have post-operative complications, whether the procedure was performed open or laparoscopically, and 2.9 times more likely to have an intra-operative complication. Moreover, frustration was significantly more likely during laparoscopic procedures (131).

These results are confounded by the difficulty of a procedure that may manifest due to existing patient-related factors that predispose him or her to a poorer outcome. On the other hand, particularly where laparoscopy is concerned, such procedural difficulty may be as much to do with simple technical and psychomotor ineptitude as it is the patient's anatomy or environmental factors. Frustration in surgery clearly requires further study, and it may be a point of further research to compare the effects of VR simulation on the levels of frustration that is experienced in the real operating environment. The suggestion is that increasing laparoscopic psychomotor skill may reduce frustration in the real OR, a possible mechanism by which surgical outcomes are improved by training on a simulator.

4.7 Trainee Motivation

An evidence-based VR curriculum may very well accelerate the acquisition of relevant skills in surgical trainees and might lead to improved patient outcomes. Even if this were not hypothetical, an otherwise-perfect curriculum would be ineffective if trainees were not sufficiently motivated to participate. Unfortunately, in the published literature it has been persistently demonstrated that attendance and participation in simulation-based curricula for the acquisition of laparoscopic skills is generally very poor, and that simply allowing unlimited access to the necessary equipment does not remedy this (133). For example, a goal-directed curriculum on a LapSim situated in the gynaecology department at a UK hospital was utilised by nine trainees for a median of only 66 minutes over a six-month period (134). Elsewhere, utilisation has been as low as 10% (135).

It is therefore of great importance to understand the factors that influence trainees' participation in VR simulation curricula, of which there are many (133). Motivators are either intrinsic or extrinsic (115); intrinsic factors are derived from the trainee him or herself, such as the desire to be a good surgeon, while extrinsic motivators originate externally, such as whether or not a curriculum is mandatory and the presence of rewards and penalties. There are factors that may influence a trainee's participation that are features of the curriculum itself, such as the presence of coaching/instruction, and the incorporated proficiency criteria for progression. A number of important factors have crystallised in the simulation literature, and are presented as follows.

4.7.1 Time and the Prioritisation of Clinical Duties

In order to progress through and successfully complete their training, trainees are required to fulfil criteria across a multitude of domains alongside their day-to-day clinical duties. In addition to learning how to operate and carry out the core task of caring for their patients, trainees must participate in audit, quality improvement, teaching and research, and so they must prioritise across these multiple competing demands (136). It is therefore of little surprise that in a systematic review exploring motivations and barriers for simulation practice in laparoscopy, lack of time was the greatest impediment to participation (133). Similarly, in the aforementioned gynaecology department, semi-structured focus group interviews revealed that time restrictions in the working schedule and prioritisation issues were the predominant barriers (134). Of 67 residents surveyed at the Yale University School of Medicine and the University of Toronto, lack of time was in the top two reasons for not participating in laparoscopy simulation training (137). Post-course surveys have shown that the time constraint is consistently cited as a reason for poor attendance (135,138,139), together with ongoing ward activities (138) and difficulties in prioritisation (134).

Many trainees also feel that laparoscopic simulation practice ought to be integrated into trainees' normal working hours (136) and that it should take place within protected time (140) if their participation is to increase. Protected time can be difficult to provide without impacting patient safety by drawing some residents away from their core responsibilities at critical moments. Alternatively, if it were left to the individual to choose the period of their schedule during which they would wish to remain undisturbed, the onus would still be on each trainee to find sufficient motivation to carve out that time. Protected time must also be considered in the context of all other competing demands for professional development, as a similar suggestion may be made if they were asked about what would motivate them across many of them and the provision of protected time for every domain may be somewhat impractical.

4.7.2 Mandatory Practice

Laparoscopic simulation practice may be encouraged if it was made a mandatory part of the surgical curriculum, though there are inconsistent findings in the literature (133). For example, 91% of the respondents to Shetty *et al*'s survey felt that it was essential for participation that laparoscopic simulation curricula were mandatory (137). 76.6% of 77 residents who responded to a survey following an at-home autonomous laparoscopic suturing course felt that such simulator training should be obligatory before operating on patients (139). Only 31% of residents who were offered a voluntary training on the ProMIS laparoscopy simulator used it at least once over a three-month period, and 64% thought that they would use it more if it were mandatory (140). In focus groups, Burden *et al* found general agreement amongst trainees that if the curriculum was mandatory then ‘they would have found more time for training’ (134). This is interesting because it serves as a reminder that ‘lack of time’ *per se* may not preclude access and participation, but there may instead be a *perception* of a lack of time when in fact the real culprit is simply not being sufficiently motivated.

Making a curriculum mandatory may not necessarily help, however. Stefanidis *et al* studied 15 PGY1 to PGY4 residents who followed a proficiency-based curriculum using FLS and VR tasks. Even though it was mandatory, weekly attendance was only 51% (141). In this example, participants were not penalised, and there are no known studies that have investigated the potential effects of penalties for non-attendance. It should not be taken for granted that making a course mandatory would yield only positive outcomes – there are concerns that this may perpetuate a tick box culture (136). Moreover, taking a ‘carrot and stick’ approach may only serve to undermine an individual’s own intrinsic motivation (115).

It is also important to consider that institutions vary in the level at which they are equipped with simulation equipment and faculty. Simulation practice requirements within a general surgical curriculum may at worst actually be impossible to fulfil, and at best could perpetuate inequality and may not result in the desired skill acquisition among all trainees. However, the ACS have to some degree set a precedent by mandating certification in the Fundamentals of Laparoscopic Surgery simulation curriculum (142), which is widely recognised. Although this does not directly mandate practice on simulators, it is difficult to comprehend that it would not have some effect on motivating residents to practice the same tasks prior to certification.

4.7.3 Goal Setting

The features of a simulation curriculum itself may profoundly impact the participation of trainees. Goal-directed learning is a core feature that leads to improved skill acquisition (94). In Stefanidis’ study of PGY1 to PGY4 residents, the impact of performance goals on motivation was rated 15 (on a scale of 1-20) in the post-course questionnaire, which correlated with the attendance rate (141).

It has previously been suggested that the ideal level at which proficiency criteria are set would be that beyond which performance in the real OR ceases to increase (94). However, it is also important to consider the impact of inappropriate goal-setting on motivation. In Burden’s study of gynaecology trainees on the LapSim, trainees reported feeling uninspired to continue with training due to continued failure to pass the later, more difficult, modules (134). Csikszentmihalyi’s pioneering discovery and characterisation of ‘flow’ teaches us about appropriate goal-setting (143). *Flow* is a state of motivation and sharp focus while performing a task, where individuals are ‘forgetting themselves in a function’ (115). It is in this state where great strides towards mastery are taken, and part of it is dependent on the appropriateness of the goals. As mentioned earlier, goals that are too easy could bore the trainee, yet too difficult and frustration and demotivation will ensue.

In the present study, many repetitions were required to move beyond the procedural LA tasks and the feedback suggested that this was indeed very frustrating for participants. However, most participants completed the curriculum (and it is not possible to conclude that the level of difficulty was responsible for the dropouts). In fact, by the time they reached the final stage, their skills had improved so significantly that very few repetitions were necessary to meet the criteria in the final task. Their perseverance may be because the trainees were often tantalisingly close to completing the module, and so it may actually be counterproductive to respond by relaxing the proficiency criteria as a means to guarding motivation. It may also be worth considering the level of effort required to progress through the curriculum and whether or not it could correlate with the intensity of subsequent feelings of reward, which was mentioned frequently in the participant diaries. Activation of the dopamine reward pathway may be relevant in explaining participants' perseverance through the curriculum (144).

'Unlocking' all stages of a curriculum (as opposed to the typical format of linear progression) has been suggested as an alternate way of maintaining motivation if trainees are often getting stuck (134). Indeed, it became clear immediately prior to the implementation part of the present study that unlocking all tasks within each stage of the curriculum was necessary to maintain motivated participants and steer them towards more deliberate practice.

4.7.4 Human Interaction

The role of coaching and instruction ('proctoring') has been discussed earlier in the context of providing feedback necessary for deliberate practice and restricting the development of bad habits that may not be recognised by the simulator. As personal trainers might play a role in motivating individuals to persist in regular physical exercise, so too might proctoring play an important role in motivating trainees' participation in simulation curricula. Yet there is very little demand for proctoring among those that have been published – in Chang's study of 29 residents on the ProMIS simulator where attendance was poor, individual proctoring was offered but there was not a single request for it (140). In another study, focus group interviews followed autonomous training of 20 residents that were randomised to training at either their homes or in a simulation centre. It was clear that the lack of direct and personalised feedback was not a concern (142). Similarly, in Burden’s study of gynaecology trainees, there was not a single request for one-to-one supervision, despite an offer being sent via email on three separate occasions, nor was any tutor approached at all during the study period (134).

These findings are markedly at odds with the regard held by participants in the present study for the coaching that was provided to them. One possible explanation is the comparative difficulty of the curricula upon which each study was based. The qualitative analysis of the present study might suggest that the curriculum’s proficiency criteria were so challenging that participants deemed coaches critical for finding the necessary performance improvements during those times when they felt their progress had stalled. And particularly during those challenging moments, the findings indicate that involvement of coaches may have had a profound effect on participants’ motivation.

Certainly, any lack of demand for proctoring may not necessarily correlate with the effect of a proctored approach on motivation and attendance, but unfortunately it has been seldom studied. Aho *et al* followed 12 Post-Graduate Year (PGY) 2 General Surgery residents who practiced on laparoscopic skills simulators both at home and in the simulator facility on campus. Each resident met with a staff surgeon at the beginning, middle and end of each rotation, where their support also included advice on how to negotiate the logistical hurdles in their attendance. In the post-test survey, more practice was reported on this rotation than on previous ones without mentor guidance, and three suggested that it could be improved by providing even more mentor-guided practice sessions (145). It was, however, noted that the concurrent service obligation within that rotation was low and the study was uncontrolled.

Elsewhere, Stefanidis *et al* described their experience with a laparoscopic simulator skills facility and reported a dramatic increase in attendance from 6% to 71% after two key interventions: they hired a skills lab coordinator and they stipulated that practice should occur within a specific hour on the first post-call day of the week (146). It is difficult to know which of the two was more significant. Clearly there is a gap in our understanding of the role of human input and guidance on trainee motivation, though the present study may have revealed some important insights.

4.7.5 Simulator Access

With time pressures and conflicting priorities, it is plausible that ease of access to the simulator equipment would have a significant influence on attendance rates. VR simulators are expensive and not very portable, and so are generally found on campus. Alternatively, there are a myriad of cheaper, portable video trainers that trainees can take home. Indeed, in Shetty *et al*'s survey, being off campus was the other of the top two reasons not to use a skills simulator (137). The hypothesis that 'home training' would result in more practice was tested by Korndorffer *et al* who randomised 20 PGY1 to PGY5 residents to either home training or centre training with video trainers and found that while the home trained group had significantly more practice sessions (13 versus 7), there was no significant difference in total practice time or number of attempts at the tasks. Those who trained in the centre reported feeling 'rushed most of the time' with a need to reorganise schedules, and they were more likely to report feeling fatigued. Fatigue and the shorter, more frequent training schedules of the home-trained group, with what is known about the effects of distributed practice on learning, may explain the trend towards better retention in this group (142). However, the results do not necessarily suggest that providing trainees with video trainers to take home would be a quick and easy way to motivate them to practice more.

4.7.6 Competition & Collaboration

Some have investigated innovative approaches to increasing trainee participation in skills simulation curricula by reminding them that their participation occurs as part of a group. Petrucci *et al* points to a major source of self-efficacy known as social modelling, where trainees witness others do the same tasks. On this basis, 14 residents were randomised either to an internet-based social collaboration tool known as 'Wiggio' or to a control group. Wiggio allows each trainee to see what their peers are doing and receive relevant notifications. There was only a trend towards better performance and attendance in the intervention group (138). This is an interesting concept, but regrettably the study was confounded by a small sample size (only six participants used the simulators) and the utilisation of a training regimen that lacked any goals for the trainees.

Wanting to win a competition has been rated by many participants as the reason for participating in a competition on the SIMENDO laparoscopic skills simulator (147). Van Dongen *et al* followed 21 surgical residents and measured participation in a VR laparoscopic skills curriculum before and after introducing a competitive element where scores were announced and a winner received a prize. The effects were characterised as 'marginal' – the number of participating residents increased from 10% to 30% and the total practice time increased from 163 to 738 minutes (135).

4.7.7 Purpose

The role of competition in motivating trainees is potentially complex since one must consider the precise reasons why it may (or may not) increase participation. For example, the desire to win may be an intrinsic motivator, or an individual may simply wish to win the prize, which would be an extrinsic motivator. This is an important distinction since, as previously highlighted and acknowledged by others (136), a 'carrot and stick' approach (i.e. rewards for attendance and punishments for non-attendance) may undermine trainees' intrinsic motivation (136) and it is quite possible that engagement would paradoxically suffer as a result. In a review of the literature in his book 'Drive', Daniel Pink demonstrates that this approach may suit simple, routine algorithmic tasks better than more complex ones, where instead motivation is cultivated when individuals are given greatest autonomy without the threat of punishment and without the change in narrative that comes with the promise of reward (115). This also calls into question the wisdom of the pervasive culture of tick-box surgical curricula (136) where trainees must score points by completing and publishing a minimum number of audit projects, research papers and other such endeavours.

In addition to autonomy, mastery and purpose have been cited as two other important components of 'true' motivation (115). The intended purpose of laparoscopic skills curricula, to improve real-world operative performance and patient outcomes by working towards skills mastery, must be relevant for trainees but this is not always the case. For example, there are some trainees who do not see the value in simulation, or at least in low fidelity simulation (136) and that can include VR (137). Relevance of simulation will be influenced by the experience of the trainee – one that is relatively proficient with LA or advanced laparoscopic suturing may not see any purpose in practicing basic psychomotor skills tasks. The effect of seniority/experience on laparoscopic skills simulation participation is seen in a small study of 10 PGY1 to PGY5 residents who underwent simulation training on the MIST-VR. PGY1-2 residents attended significantly more training sessions than did the PGY3-5 residents, though there may have been a multitude of confounding factors (148). Nevertheless, this finding is echoed more broadly in Gostlow *et al*'s systematic review of influencers and barriers to laparoscopic skills simulation training, and it is suggested that this relative lack of purpose for more senior trainees may be responsible for some of the poor attendance seen in surgical skills curricula (133,138).

Fidelity may need to be tailored to the level of trainees' experience, as it has been suggested that junior trainees may prefer video box trainers and VR, whereas more experienced trainees prefer live animal and explanted tissue models (137). The focus groups in Blackhall *et al*'s study expressed a preference for laparoscopic procedural modules over general psychomotor skills tasks as it is easier to appreciate their relevance (136).

Ultimately, it may be important to ensure that trainees understand the value of surgical skills simulation, why it is relevant to each individual and what impact it could have. From our understanding of motivation, it may be prudent to highlight those aspects that chime with their intrinsic motivation such as helping them appreciate that their own real-world performance will improve, that the outcomes for their patients may improve, and that it may afford them greater autonomy in the real OR. This may achieve more than highlighting any extrinsic motivating factors such as 'because you have to'. Of course, this entire hypothesis has yet to be tested.

A perceived lack of a link between VR simulation practice and real-world clinical operating (134) might affect a trainee's sense of purpose. In the UK, some trainees also have the perception that there is little point in laparoscopic skills simulation practice when they are either too busy to put it into practice, or where there is too little opportunity for such real-world experience (for example, if the trainee is in a plastic surgery rotation) (136). It has therefore been suggested that such simulation should be coupled with clinical practice (136). Engaging trainers might be crucial – if they see the value in laparoscopic skills simulation and its transferability to the real OR, then coupling their experience with the real world may be better facilitated.

There have been great strides made in the development of technology and simulation-based curricula for the acquisition of laparoscopic skills by trainees. On the other hand, trainee engagement is generally very low and the quality of the studies that investigate this aspect are generally very poor. Gostlow *et al*'s systematic review highlights this – only nine papers were included, many of which had a very small sample size in single institutions with high dropout rates (133). Studies are also often uncontrolled and with factors that confound them, such as the utilisation of aimless rather than goal-directed practice, and healthcare and training systems vary widely between the US and UK, which further limit the applicability of the published evidence. There is a great need for high quality randomised studies, to isolate the effects of a single intervention on trainee motivation and participation in SBME. Specifically, it may be worthwhile concentrating efforts to investigate the effects of extrinsic versus intrinsic motivators.

In comparison to the above-cited literature, trainee engagement in the present study was comparably very high with a low drop-out rate, but it would be worth emphasising that this does not imply that the curriculum and associated infrastructure led to high levels motivation. The denominator of many of the aforementioned studies were *all* residents within a particular institution, while that of the present study were merely all recipients of the email invitations to participate. The reasons for the relatively low number of dropouts were not provided, but even if they were all to have cited the level of difficulty as the reason, motivation levels could still be reasonably considered as very high and so it would be worth considering the relevant factors.

Participation in the curriculum was not mandatory, and trainees will have experienced conflicting priorities and time pressures. The CSBS's 24-hour open access policy and its location 'on campus' may have helped to alleviate this. However, many trainees came from other hospitals. Furthermore, contrary to the belief by some that trainees would not attend during their time off (140), our experience was that this was indeed occurring. Trainees also knew that upon completion of the curriculum, they would be given a certificate as an external reward. The curriculum was also clearly goal-directed with distinct phases of mandatory proctoring, in addition to their being a full-time facilitator managing the centre, who was also laparoscopically and educationally trained. It is also worth considering that trainees signed a 'contract' to indicate their agreement to the rules, and this may have behaved as a 'contract of commitment' (145) though its effect on trainee motivation is yet to be investigated.

With the importance of 'purpose' in mind, an important factor may have been that the curriculum was specifically aimed at trainees who were both novice laparoscopic surgeons whose participation was most likely to be coupled to real-world experiences near the beginning of their careers. Clearly there are many intertwining factors at play, but at very least being able to identify them is a pre-requisite for future research that is greatly needed in the field of trainee motivation and participation in SBME.

4.8 Faculty Recruitment & Retention

Educational endeavours that involve VR simulation rely on faculty that have an important role in developing curricula, teaching, assessing performance and developing research programmes (149). With specific regards to surgical skill, having a skilled mentor is very important, according to the President of the Southern Surgeons Association (150), though finding and retaining faculty with the appropriate skills in simulation is a perennial difficulty (151). Indeed, simulation is time-consuming and faculty-intensive (151) and the VR LA curriculum itself depends on faculty to deliver simulator induction, instruction, coaching and assessment.

Regarding the difficulty in faculty recruitment and retention, the finger has been pointed towards relatively new working time regulations, which exist in both Europe and the US. In a systematic review of studies that have investigated the effects of such restrictions in the US, 20 examined the effects on surgical faculty. The main findings were that they often have less time for non-clinical activities such as teaching, as they are frequently required to pick up the slack in clinical duties due to residents no longer having enough time to complete their usual work. In addition to the work shifting from residents to faculty, other effects included less time for teaching, worsening quality of life and more job dissatisfaction (152). Faculty attrition in the US has also been blamed on physician burnout, lack of mentorship, and work-life balance difficulties (153).

Working time restrictions in the UK have resulted in frequent rota gaps, with clinical commitments making it extremely challenging to recruit faculty (154). Surveys and site visits carried out in a teaching deanery in the southeast of England revealed four main problems in maintaining adequate numbers of faculty for foundation doctor simulation: 1) a reliance on voluntary faculty, 2) permanent faculty being too busy to find volunteer faculty, 3) a lack of interest in simulation outside anaesthesia and medicine, and 4) a difficulty in clinical staff being released from clinical duties (155). The last point suggests that ‘protected time’ may play an important role in solving the problem especially since teaching has generally not been recognised in UK job plans (151).

Such protected time would inevitably require dedicated funding, in addition to that needed to cover costs for any simulation centre, such as rental fees, other full-time staff, supplies, materials, travel and contract services (for example, simulator maintenance) (149). VR simulators are particularly expensive, with costs that can approach £120,000 GBP ($197,000 USD) (156). Yet in the US, 15% of education institutes accredited by the ACS receive no funding from any medical school or hospital (150).

Kim *et al* described how innovation in the way educational services are organised and funded can lead to improved recruitment and retention of surgical faculty. At the University of Washington, clinical departments pay part of the salary of faculty at the Institute for Simulation and Interprofessional Studies, and in return their residents are free to use their facilities and benefit from their educational output. Clinical departments also receive reports on their faculty’s performance. They found that protected time was easier to obtain for faculty when there is formal affiliation between the educational and clinical departments. Faculty was increased by 500% with excellent retention rates and many logged teaching hours for the facility. Emphasis was placed on the importance of aligning simulation goals with the hospital's mission, and they discuss previous findings that highlight mentoring and the establishment of networks to improve faculty satisfaction, productivity, institutional loyalty and retention (149).

At an NHS Trust in the UK, Arora *et al* described how Associate Simulation Fellows were successfully recruited and trained with no additional financial burden to the hospital Trust, though it is not clear how this was achieved (154). The notable difference, however, between the UK and US is that trainees in the former frequently move on as their training programme rotates them at least once per year. Rotations in the UK can cover wide geographical regions, which could make it particularly challenging to retain faculty once they rotate. In the US, residents can often expect to stay in their hospital for the duration of their programme.

Financial independence allows resources to be directed more creatively and effectively (153), and managing a simulation centre as a profitable enterprise may help to generate the funding necessary to recruit and retain faculty. The CSBS is exceptional in that it manages to both develop and implement surgical skills curricula. Its central location in London makes it accessible to trainees across three separate training programmes, and a trainee rotating to a different hospital does not necessarily mean a termination of his or her involvement as faculty or participant in skills curricula at this centre. But training must be equitable across the UK without regional variation; such a facility must be the rule rather than the exception, and this begins with setting national standards for simulation training.

The Association for Simulated Practice in Healthcare (ASPiH) is the UK national body that focuses on SBME and technology-enhanced learning (TEL). It was part-funded by Health Education England (HEE) to produce an evidence-based framework of standards. ASPiH recognises the importance of support from deans and hospital leads in enabling the dedication of time and financial support for a facility's development. They recommend 'a designated lead with organisational influence and accountability' to manage simulation activity, in addition to emphasising the importance of the recruitment and retention of simulation faculty. SBME facilities must include individuals with the appropriate technological expertise to support a simulation programme, and faculty who are experts in the procedures taught, and who have had specific simulation equipment training prior to independently facilitating a course in procedural simulation. For such faculty, protected time is recommended, with engagement of healthcare organisations and educational institutions to incorporate their role as faculty within contracts, job plans and linking it to appraisal. Novice faculty should be mentored and all faculty should be appropriately recognised in order to maximise retention (157). ASPiH also suggested the consideration of 'Simulation Fellowship Programmes' such as that mentioned earlier, adopted by West Hertfordshire Hospitals NHS Trust (154).

It is worth noting that the heavy focus of this framework on the delivery of SBME is in non-surgical settings and 'scenario-based simulation' – most of the standards, including those under the 'faculty' section, have been developed in this context. Debriefing is said to be the most important component of SBME (157), yet it is perhaps not as relevant in the SBME of surgical psychomotor skills. In such a clearly-defined domain, training facilitators may not need to be so heavily involved in all facets of traditional simulation faculty. The VR LA curriculum works well, partly because the participating faculty have very precise involvement – to assess performance on a simulator and provide operative feedback. They need not be involved in research and curriculum development, and so the magnitude of any commitment should not deter them if their interest lies in only one part of that traditional role. On the other hand, the nature of distributed practice presents unique challenges to faculty, since they are required frequently, for very short periods, and at unpredictable times. This would be particularly difficult for those based in a separate geographical location, where such 'limited scope' faculty may find it difficult to make a special trip across a city to deliver only 15 minutes of teaching.

Recommendations specific to SBME in the acquisition of surgical skills have been made by The Royal College of Surgeons in their publication entitled *Improving Surgical Training*, recommending profound changes to the structure of surgical training on the basis that there is currently too much emphasis on service delivery. Recommendation 14 states that 'simulation should be embedded and enhanced within the surgical curricula and there should be sufficient resource to ensure availability to all trainees', and recommendation 15 states that 'each phase of training should be preceded by an educational induction where technical and non-technical skills are taught and developed in a simulated environment'. It was therefore planned for the JCST to approach the General Medical Council (GMC) to embed simulation as a mandatory component of Core Surgical Training. This would clearly depend on the recruitment and retention of faculty and so concerns about funding of simulation training, in addition to equity of access, were duly expressed (158). Unfortunately, there seems to have been little progress, which may reflect the wide gap between the ease at which recommendations can be made versus the ease at which they can be fulfilled. There is much to be learned and applied from the innovative approaches taken by the University of Washington and the Royal Free Hospital's CSBS.

4.9 Study Limitations & Future Research

The curriculum in this present study has been developed upon a strong theoretical framework. Proficiency criteria were derived from Part 1, which used a methodology that mitigated against warm-up effects through the randomisation of both experts and novices. Nevertheless, the manner in which tasks and metrics were selected for inclusion in the curriculum presents a potential area of weakness in our methodology, since it was subject only to internal consensus and scrutiny. A more robust method to draw from both internal and external expertise could have been utilised (for example, the Delphi technique (159)). However, this is likely to have been futile since there are so few surgeons with LAP Mentor experience. Instead, Table 14, Table 15 and Table 16 were constructed in order to provide full transparency regarding the rationale to include or exclude individual metrics.

Also worth noting is the changeover in hardware that occurred during the study. It began using the LAP Mentor II, but the haptics technology was updated following Part 1, and participants were using the LAP Mentor III for Part 3 (implementation), which may have affected performances. Since this changeover occurred between and not during individual parts of the study, and the update affected only the haptics, the authors do not believe it will have significantly confounded any of the findings subsequently. Updating technology is important, not just to further increase simulator fidelity, but to iron out glitches which we have demonstrated to occur frequently enough to be a considerable source of frustration and can themselves confound the results of studies designed to assess their validity.

Whilst our curriculum may ostensibly teach the psychomotor skills and the key steps in performing LA, appendicitis is an acute inflammatory condition that arguably lacks the anatomical predictability of an elective cholecystectomy or an emergency salpingectomy. The intraoperative presentation of acute appendicitis varies in the position of the appendix and the degree of inflammation and adhesions. In reality, a surgeon must be able to perform a wide array of technical steps for each appendicectomy as they become necessary, such as caecal mobilisation and adhesiolysis, in addition to the cognitive aspects of recognising the appendix base and judging the securest method of ligation. None of these steps are included in the LA module and it may present a weakness of the curriculum. Furthermore, it does not address other important domains in the OR, such as knowledge of anatomy and decision-making with task-switching in case of unexpected events (160), although we would expect that the LA curriculum would complement these cognitive domains as part of the wider curriculum for surgical training.

Rates of participant dropout in Part 1 was very low. However, some dropouts occurred in Part 3 and the reasons for this remain unknown. Given that the curriculum was so challenging, this might limit our understanding of its real-world validity. Part 3 also occurred during a once-in-a-century event – the Covid-19 pandemic. Participants' clinical duties and shift patterns will have been extraordinarily distorted due to redeployment to cope with the demand of Covid-19 waves. This inevitably led to some gaps in continuity of training and so may have affected learning curves. It is also conceivable that levels of fatigue and focus on extracurricular training were inadvertently affected. This unique context might potentially affect any comparisons between the findings of the present study and those of others.

The total number of participants in Part 3 was small, rendering it impossible to make further conclusions about the effects of prolonged practice, particularly on performance variability. However, Part 3 was primarily qualitative in nature and therefore sufficient for meeting the study's primary and secondary objectives (the curriculum was implemented and data saturation was confirmed using a validated method). After deliberation, it was decided that performance data from the dropouts in Part 3 should be excluded from learning curves. This was in order to reduce heterogeneity, as this may have confused interpretation of the results.

The simulator itself posed unique challenges to the study and curriculum. There were frequent software crashes which led to missing performance data that was not included in the analysis. It is unclear how unrecorded practice may have affected the results. Other software glitches were commented upon in participants diaries, which suggested that at least some of the curriculum saw participants learning how to 'beat' the simulator, rather than become competent laparoscopic surgeons. This was limited to one small component of the procedural tasks, and addressing it should require a software update.

The software controlling the haptic feedback had actually been updated after the LA curriculum was constructed the study but prior to its implementation, which may have potentially confounded results. Since the effect of haptic feedback on training is probably only mild, it is unlikely to have made any meaningful impact. However, faculty were informally reporting a noticeable change in the quality of haptic feedback during PT1 and it was felt that there was a considerable risk that this may have had a deleterious effect on performance for this single task.

Such hardware and software updates may indeed have the potential to limit the future validity of the data upon which the curriculum is based, as the simulator and appendicectomy module continue to evolve. Theoretically, the module may eventually hold only a loose semblance to the version that has been validated. Frequent redesigns of the curriculum based upon new data is probably not feasible, so it is important that manufacturers involve stakeholders (including educationalists) closely when new updates are made. In any case, it is difficult to imagine any changes that would be radical enough to compromise any real-world gains in performance when trainees follow a curriculum that was designed using earlier versions of the module or simulator.

The curriculum itself requires significant time and commitment from participants, and it had been implemented inside a single centre of one of the largest cities in the world. This study has demonstrated that many trainees are sufficiently motivated to travel frequently and commit themselves to a long course of training. On the other hand, there are many more who will not have responded to invitations to participate. The value of this key denominator is uncertain, and although this can be addressed in future rounds of recruitment for real-world training, a baseline value is useful in order to measure the effects of key interventions on the motivation of trainees to participate in the LA curriculum.

It would seem that trainee motivation forms one of two key areas of necessary future research. The curriculum is founded on a regimen of distributed practice, and so by its nature requires commitment that is sustained over a relatively prolonged period of time. The CSBS is a single unit in a very large city, and so it would be crucial to understand the factors that would encourage trainees to travel regularly in order to complete the curriculum.

Participant diaries have suggested that the curriculum is very challenging and often frustrating. This is one of very few studies that demonstrate the important role of instructors in trainee motivation and the mitigation against frustration in SBME. Thus, further qualitative research would be invaluable to corroborate these findings and build on the evidence-base that will influence future SBME curricula.

A sense of purpose is a key ingredient for motivation and reaching states of flow. In an attempt to enhance this, the curriculum was aimed not only at novice laparoscopists, but specifically those who may be most likely to be in a position to see their skills tested in the real operating environment. Moving forward, it would be useful to study the effects of purpose, and of tailoring the timing of the curriculum’s deployment, on trainee motivation.

Consideration should be given to where and how such a VR curriculum should lie in the context of wider surgical training and the non-technical aspects of LA, such as decision-making and perioperative care. The potentially-mixed effects and drawbacks of making any curriculum mandatory have been discussed, and it has been determined that much of the novice and expert cohort of the present study would support this. Broader trials within the wider training programme would be necessary to assess the impact of such a change, but this may first require robust data demonstrating the translation of skills, which is the second of two key areas of future research. Demonstrating positive effects of our appendicectomy curriculum on real-world operative performance and patient outcomes would be difficult for training providers to ignore, particularly since most are performed by trainees and appendicitis is so very common.

There are other gaps in the collective understanding as regards to building a curriculum for laparoscopic skills training. For example, the optimal practice session duration is still uncertain, although we believe 45 minutes is likely to be the upper limit, particularly in view of the participant testimony presented in the present study. The optimal duration of spacing intervals between sessions is also unclear, as is the cost-benefit ratio of haptic feedback in VR laparoscopy simulators. The effect of cognitive loading on surgical performance is an emerging area of research and education innovation that is highly relevant, given the discrepancies between real-world operating pressures and the sterile environment of a simulation suite.

The authors' plans for future work primarily focuses on enrolling more trainees into the LA curriculum, to analyse their participant diaries and gain detailed feedback in order to make further improvements that would positively impact trainee involvement, satisfaction and motivation. There are also plans to design a research protocol to test the hypothesis that completion of the LA curriculum translates to better performance in the real OR (predictive validity).

5 Conclusion

This study has met its primary and secondary objectives: construct validity of several BS tasks and the LA module of the LAP Mentor was demonstrated. A structured curriculum was produced upon a strong, evidence-based theoretical framework. The curriculum was implemented among novice trainees, who demonstrated performance improvements.

Finally, this implementation was subsequently evaluated. The analysis of participant diaries revealed that the curriculum was a source of both reward and frustration. The latter resulted from the challenging nature of the curriculum and partly due to technical failures of the simulator.

Strong support was found in favour of the inclusion of human-delivered feedback sessions, which appears to have facilitated deliberate practice, participant motivation and the mitigation against the adoption of unsafe surgical techniques.

The value of haptic feedback in the context of the curriculum is unproven. The authors believe its value lies in improving the fidelity of the machine, which may positively impact trainee motivation and participation, rather than directly on the acquisition of psychomotor skills.

The open access set-up of the CSBS allowed participants flexibility to practice around their individual schedules. However, this occasionally led to clashes and distraction when participants arrived to find the simulators already in use.

Following the successful implementation of the LA curriculum, priority should now be given to optimising the curriculum for trainee participation and determining its effect on the real-world performance of LA.

6 Bibliography

1. Chowdhury MM, Dagash H, Pierro A. A systematic review of the impact of volume of surgery and specialization on patient outcome. Br J Surg. 2007;94(2):145–61.
2. Allum B. Chairman’s Update. JCST Newsl. 2015;1–7.
3. Cook T, Lund J. General Surgery Curriculum. Intercoll Surg Curric Program. 2021;1–97.
4. Bhangu A, Richardson C, Torrance A, Pinkney T, Battersby C, Beral D, et al. Multicentre observational study of performance variation in provision and outcome of emergency appendicectomy. Br J Surg. 2013;100(9):1240–52.
5. Madden GBP, Madden AP. Has Modernising Medical Careers lost its way? BMJ. 2007;335(7617):426–8.
6. Mehmood S, Anwar S, Ahmed J, Tayyab M, O’Regan D. A survey of UK surgical trainees and trainers; latest reforms well understood but perceived detrimental to surgical training. Surg. 2012;10(1):9–15.
7. Collum J, Harrop J, Stokes M, Kendall D. Patient safety and quality of care continue to improve in NHS North West following early implementation of the European Working Time Directive. QJM. 2010;103(12):929–40.
8. Cappuccio FP, Bakewell A, Taggart FM, Ward G, Ji C, Sullivan JP, et al. Implementing a 48 h EWTD-compliant rota for junior doctors in the UK does not compromise patients’ safety: Assessor-blind pilot comparison. QJM. 2009;102(4):271–82.
9. Datta ST, Davies SJ. Training for the future NHS: Training junior doctors in the United Kingdom within the 48-hour European working time directive. BMC Med Educ. 2014;14 Suppl 1(Suppl 1):S12.
10. NCEPOD. Summary of the 1995/96 Report “Who operates when?”. 1997.
11. Jackson GP, Tarpley J. How long does it take to train a surgeon? BMJ. 2009;339(7729):1062–4.
12. Lambert TW, Smith F, Goldacre MJ. The impact of the European Working Time

Directive 10 years on: views of the UK medical graduates of 2002 surveyed in 2013–2014. JRSM Open. 2016;7(3):1–8.

13. Navez B, Navez J. Laparoscopy in the acute abdomen. Best Pract Res Clin Gastroenterol. 2014;28(1):3–17.

14. Munz Y, Kumar BD, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers: Is one superior to the other? Surg Endosc Other Interv Tech. 2004;18(3):485–94.

15. Lehmann KS, Ritz JP, Maass H, Çakmak HK, Kuehnapfel UG, Germer CT, et al. A prospective randomized study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. Ann Surg. 2005;241(3):442–9.

16. Dent TL. The impact of laparoscopic surgery on health care delivery. The learning curve: skills and privileges. J Laparoendosc Surg. 1993;3(3):247–9.

17. Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. N Engl J Med. 1991;324:1073–8.

18. Deziel DJ, Millikan KW, Economou SG, Doolas A, Ko ST, Airan MC. Complications of laparoscopic cholecystectomy: A national survey of 4,292 hospitals and an analysis of 77,604 cases. Am J Surg. 1993;165(1):9–14.

19. Gobet F, Lane PCR, Croker S, Cheng PCH, Jones G, Oliver I, et al. Chunking mechanisms in human learning. Trends Cogn Sci. 2001;5(6):236–43.

20. Baird DLH, Simillis C, Kontovounisios C, Rasheed S, Tekkis PP. Acute appendicitis. BMJ. 2017;357:j1703.

21. Hospital Episode Statistics for England. Admitted Patient Care statistics, 2018-2019. Heal Soc Care Inf Cent. 2019.

22. Varadhan KK, Neal KR, Lobo DN. Safety and efficacy of antibiotics compared with appendicectomy for treatment of uncomplicated acute appendicitis: meta-analysis of randomised controlled trials. BMJ. 2012;344(e2156).

23. Semm K. Endoscopic Appendectomy. Endoscopy. 1983;15(02):59–64.

24. Jaschinski T, Mosch CG, Eikermann M, Neugebauer EA, Sauerland S. Laparoscopic versus open surgery for suspected appendicitis. Cochrane Database Syst Rev. 2018;11(11):CD001546.

25. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ. 2009;338(may14 2):b1802–b1802.

26. Mahajan RP. The WHO surgical checklist. Best Pract Res Clin Anaesthesiol. 2011;25(2):161–8.

27. The Royal College of Surgeons of England. Core Skills in Laparoscopic Surgery. Available from: https://www.rcseng.ac.uk/education-and-exams/courses/search/core-skills-in-laparoscopic-surgery/

28. The Royal College of Surgeons of England. Intermediate Skills in Laparoscopic Surgery. Available from: https://www.rcseng.ac.uk/education-and-exams/courses/search/intermediate-skills-in-laparoscopic-surgery/

29. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev. 1993;100(3):363–406.

30. Ericsson KA, Nandagopal K, Roring RW. Toward a science of exceptional achievement: Attaining superior performance through deliberate practice. In: Annals of the New York Academy of Sciences. 2009. p. 199–217.

31. Fitts PM, Posner MI. Human Performance. Basic Concepts in Psychology. Brooks/Cole Publishing Co. 1967.

32. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: The relationship between clinical experience and quality of health care. Ann Intern Med. 2005;142(4):260–73.

33. Ericsson KA. The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance. In: The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press. 2006.

34. Simon HA, Chase WG. Skill in chess. Am Sci. 1973;61(4):394–403.

35. Helsen WF, Hodges NJ, Winckel J Van, Starkes JL. The roles of talent, physical precocity and practice in the development of soccer expertise. J Sports Sci.

2000;18(9):727–36.

36. Udani AD, Macario A, Nandagopal K, Tanaka MA, Tanaka PP. Simulation-Based Mastery Learning with Deliberate Practice Improves Clinical Performance in Spinal Anesthesia. Anesthesiol Res Pract. 2014;2014:659160.

37. Barsuk JH, McGaghie WC, Cohen ER, Balachandran JS, Wayne DB. Use of simulation-based mastery learning to improve the quality of central venous catheter placement in a medical intensive care unit. J Hosp Med. 2009;4(7):397–403.

38. Mavroudis CD, Mavroudis C, Jacobs JP, DeCampli WM, Tweddell JS. Simulation and Deliberate Practice in a Porcine Model for Congenital Heart Surgery Training. Ann Thorac Surg. 2018;105:637–44.

39. Nesbitt JC, St Julien J, Absi TS, Ahmad RM, Grogan EL, Balaguer JM, et al. Tissue-based coronary surgery simulation: Medical student deliberate practice can achieve equivalency to senior surgery residents. J Thorac Cardiovasc Surg. 2013;145(6):1453–9.

40. Udani AD, Harrison TK, Mariano ER, Derby R, Kan J, Ganaway T, et al. Comparative-effectiveness of simulation-based deliberate practice versus self-guided practice on resident anesthesiologists' acquisition of ultrasound-guided regional anesthesia skills. Reg Anesth Pain Med. 2016;41(2):151–7.

41. Snyder CW, Vandromme MJ, Tyra SL, Hawn MT. Retention of colonoscopy skills after virtual reality simulator training by independent and proctored methods. Am Surg. 2010;76(7):743–6.

42. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med. 2011;86(6):706–11.

43. Palter VN, Grantcharov TP. Individualized deliberate practice on a virtual reality simulator improves technical performance of surgical novices in the operating room: A randomized controlled trial. Ann Surg. 2014;259(3):443–8.

44. Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, Grantcharov T, et al.

Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. Ann Surg. 2011;253(6):1216–22.

45. Hashimoto DA, Sirimanna P, Gomez ED, Beyer-Berjot L, Ericsson KA, Williams NN, et al. Deliberate practice enhances quality of laparoscopic surgical performance in a randomized controlled trial: from arrested development to expert performance. Surg Endosc Other Interv Tech. 2015;29(11):3154–62.

46. Boyle E, Al-Akash M, Gallagher AG, Traynor O, Hill ADK, Neary PC. Optimising surgical training: Use of feedback to reduce errors during a simulated surgical procedure. Vol. 87, Postgraduate Medical Journal. 2011. p. 524–8.

47. Buescher JF, Mehdorn AS, Neumann PA, Becker F, Eichelmann AK, Pankratius U, et al. Effect of Continuous Motion Parameter Feedback on Laparoscopic Simulation Training: A Prospective Randomized Controlled Trial on Skill Acquisition and Retention. Journal of Surgical Education. 2017.

48. Snyder CW, Vandromme MJ, Tyra SL, Hawn MT. Proficiency-Based Laparoscopic and Endoscopic Training With Virtual Reality Simulators: A Comparison of Proctored and Independent Approaches. J Surg Educ. 2009;66(4):201–7.

49. Shippey SH, Chen TL, Chou B, Knoepp LR, Bowen CW, Handa VL. Teaching subcuticular suturing to medical students: Video versus expert instructor feedback. J Surg Educ. 2011;68(5):397–402.

50. Nousiainen M, Brydges R, Backstein D, Dubrowski A. Comparison of expert instruction and computer-based video training in teaching fundamental surgical skills to medical students. Surgery. 2008;143(4):539–44.

51. Bjerrum F, Maagaard M, Led Sorensen J, Rifbjerg Larsen C, Ringsted C, Winkel P, et al. Effect of Instructor feedback on skills retention after laparoscopic simulator training: Follow-up of a randomized trial. J Surg Educ. 2015;72(1):53–60.

52. Hatala R, Cook DA, Zendejas B, Hamstra SJ, Brydges R. Feedback for simulation-based procedural skills training: A meta-analysis and critical narrative synthesis. Vol. 19, Advances in Health Sciences Education. 2014. p. 251–72.

53. Schmidt R a, Lee TD. Motor control and learning: A behavioral emphasis (5th ed.).

Human Kinetics. 2011;581–ix, 581.

54. Wulf G, Shea CH. Understanding the role of augmented feedback : The good, the bad and the ugly. In: Skill Acquisition in Sport. Routledge. 2004. p. 145–68.

55. Boyle E, O'Keeffe DA, Naughton PA, Hill ADK, McDonnell CO, Moneley D. The importance of expert feedback during endovascular simulator training. J Vasc Surg. 2011;54(1).

56. Strandbygaard J, Bjerrum F, Maagaard M, Winkel P, Larsen CR, Ringsted C, et al. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: A randomized trial. Ann Surg. 2013;257(5):839–44.

57. Benjamin AS, Tullis J. What makes distributed practice effective? Cogn Psychol. 2010;61(3):228–47.

58. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychol Bull. 2006;132(3):354–80.

59. Honda M, Deiber MP, Ibáñez V, Pascual-Leone A, Zhuang P, Hallett M. Dynamic cortical involvement in implicit and explicit motor sequence learning. A PET study. Brain. 1998;121(11):2159–73.

60. Mackay S, Morgan P, Datta V, Chang A, Darzi A. Practice distribution in procedural skills training: A randomized controlled trial. Surg Endosc Other Interv Tech. 2002;16(6):957–61.

61. Moulton CAE, Dubrowski A, MacRae H, Graham B, Grober E, Reznick R. Teaching surgical skills: What kind of practice makes perfect? A randomized, controlled trial. Ann Surg. 2006;244:400–9.

62. Cecilio-Fernandes D, Cnossen F, Jaarsma DADC, Tio RA. Avoiding Surgical Skill Decay: A Systematic Review on the Spacing of Training Sessions. J Surg Educ. 2018;75(2):471–80.

63. Spruit EN, Band GPH, Hamming JF. Increasing efficiency of surgical training: effects of spacing practice on skill acquisition and retention in laparoscopy training. Surg Endosc. 2015;29(8):2235–43.

64. Gallagher AG, Jordan-Black JA, O'Sullivan GC. Prospective, randomized assessment of

the acquisition, maintenance, and loss of laparoscopic skills. Ann Surg. 2012;256(2):387–93.

65. Schoeff S, Hernandez B, Robinson DJ, Jameson MJ, Shonka DC. Microvascular anastomosis simulation using a chicken thigh model: Interval versus massed training. Laryngoscope. 2017;127(11):2490–4.

66. Brashers-Krug T, Shadmehr R, Bizzi E. Consolidation in human motor memory. Nature. 1996;382(6588):252–5.

67. Kuhn M, Wolf E, Maier JG, Mainberger F, Feige B, Schmid H, et al. Sleep recalibrates homeostatic and associative synaptic plasticity in the human cortex. Nat Commun. 2016;7.

68. Donovan JJ, Radosevich DJ. A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. J Appl Psychol. 1999;84:795–805.

69. De Win G, Van Bruwaene S, De Ridder D, Miserez M. The optimal frequency of endoscopic skill labs for training and skill retention on suturing: A randomized controlled trial. J Surg Educ. 2013;70(3):384–93.

70. Mitchell EL, Lee DY, Sevdalis N, Partsafas AW, Landry GJ, Liem TK, et al. Evaluation of distributed practice schedules on retention of a newly acquired surgical skill: A randomized trial. Am J Surg. 2011;201(1):31–9.

71. Bjerrum AS, Eika B, Charles P, Hilberg O. Distributed practice. The more the merrier? A randomised bronchoscopy simulation study. Med Educ Online. 2016;21(1).

72. Available from: https://london.hee.nhs.uk/specialty-schools/surgery

73. Chmarra MK, Dankelman J, van den Dobbelsteen JJ, Jansen F-W. Force feedback and basic laparoscopic skills. Surg Endosc. 2008 Apr 29;22:2140–8.

74. Zhou M, Tse S, Derevianko A, Jones DB, Schwaitzberg SD, Cao CGL. Effect of haptic feedback in laparoscopic surgery skill acquisition. Surg Endosc. 2012;26(4):1128–34.

75. Wagner CR, Stylopoulos N, Howe RD. The Role Of Force Feedback In Surgery: Analysis Of Blunt Dissection. Haptic Interfaces for Virtual Environment and Teleoperator Systems, International Symposium on. 2002. 73–73 p.

76. Hagelsteen K, Johansson R, Ekelund M, Bergenfelz A, Anderberg M. Performance and perception of haptic feedback in a laparoscopic 3D virtual reality simulator. Minim Invasive Ther Allied Technol. 2019 Sep 3;28(5):309–16.

77. Panait L, Akkary E, Bell RL, Roberts KE, Dudrick SJ, Duffy AJ. The Role of Haptic Feedback in Laparoscopic Simulation Training. J Surg Res. 2009 Oct 1;156(2):312–6.

78. Thompson JR, Leonard AC, Doarn CR, Roesch MJ, Broderick TJ. Limited value of haptics in virtual reality laparoscopic cholecystectomy training. Surg Endosc. 2011 Sep 25;25:1107–14.

79. Overtoom EM, Horeman T, Jansen FW, Dankelman J, Schreuder HWR. Haptic Feedback, Force Feedback, and Force-Sensing in Simulation Training for Laparoscopy: A Systematic Overview. J Surg Educ. 2018 Jan 1;76(1):242–61.

80. Picod G, Jambon A, Vinatier D, Dubois P. What can the operator actually feel when performing a laparoscopy? Surg Endosc. 2005;19:95–100.

81. Trejos AL, Patel R V., Malthaner RA, Schlachta CM. Development of force-based metrics for skills assessment in minimally invasive surgery. Surg Endosc. 2014 Feb 12;28:2106–19.

82. American Psychological Association Association, American Educational Research Education, National Council on Measurements in Education. Technical recommendations for psychological tests and diagnostic techniques. Psychol Bull. 1954;21(2 (Suppl)).

83. American Psychological Association Association, American Educational Research Education, National Council on Measurements in Education. Standards for Educational and Psychological Tests and Manuals. Washington, D.C.: American Psychological Association.;

84. Ruch GM. The improvement of the written examination. Chicago: Scott, Foreman and Company.; 1924.

85. Available from: http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2013/06/2013-Meaning-of-validity-Oxford-v4-slides.pdf

86. Messick, S. (1989). Validity. In R. Linn (Ed.). Educational Measurement (3rd edition)

(pp.13–100). Washington, DC: American Council on Education.

87. American Educational Research Association, American Psychological Association NC on M in E. Standards for Educational and Psychological Testing. Washington, D.C.: American Educational Research Association.; 1999.

88. Borgersen NJ, Naur TMH, Sørensen SMD, Bjerrum F, Konge L, Subhi Y, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. Ann Surg. 2018 Jun 1;267(6):1063–8.

89. Korndorffer JR, Kasten SJ, Downing SM. A call for the utilization of consensus standards in the surgical education literature. Am J Surg. 2010 Jan;199(1):99–104.

90. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R, Cook DA, et al. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Heal Sci Educ. 2014;19:233–50.

91. Sirimanna P, Gladman MA. Development of a proficiency-based virtual reality simulation training curriculum for laparoscopic appendicectomy. ANZ J Surg. 2017;87(10):760–6.

92. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery. 2004 Jan;135(1):21–7.

93. http://www.rfh-simulator-centre.co.uk

94. Sinitsky DM, Fernando B, Berlingieri P. Establishing a curriculum for the acquisition of laparoscopic psychomotor skills in the virtual reality environment. Am J Surg. 2012;204(3):367–76.

95. Aggarwal R, Grantcharov TP, Eriksen JR, Blirup D, Kristiansen VB, Funch-Jensen P, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg. 2006;244(2):310–4.

96. Aggarwal R, Crochet P, Dias A, Misra A, Ziprin P, Darzi A. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. Br J Surg. 2009;96(9):1086–93.

97. Brewin J, Nedas T, Challacombe B, Elhage O, Keisu J, Dasgupta P. Face, content and construct validation of the first virtual reality laparoscopic nephrectomy simulator.

BJU Int. 2010;106(6):850–4.

98. Bharathan R, Vali S, Setchell T, Miskry T, Darzi A, Aggarwal R. Psychomotor skills and cognitive load training on a virtual reality laparoscopic simulator for tubal surgery is effective. Eur J Obstet Gynecol Reprod Biol. 2013;169(2):347–52.

99. Wilson M, McGrath J, Vine S, Brewer J, Defriend D, Masters R. Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts. Surg Endosc Other Interv Tech. 2010;24(10):2458–64.

100. Yamaguchi S, Konishi K, Yasunaga T, Yoshida D, Kinjo N, Kobayashi K, et al. Construct validity for eye-hand coordination skill on a virtual reality laparoscopic surgical simulator. Surg Endosc. 2007;21(12):2253–7.

101. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. Am J Surg. 2003;185(2):146–9.

102. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. Br J Surg. 2004;911549–58.

103. Larsen CR, Grantcharov T, Aggarwal R, Tully a, Sørensen JL, Dalsgaard T, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc. 2006;20(9):1460–6.

104. Sinitsky DM, Fernando B, Potts H, Lykoudis P, Hamilton G, Berlingieri P. Development of a structured virtual reality curriculum for laparoscopic appendicectomy. Am J Surg. 2020;219(4):613–21.

105. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2005;190(1):107–13.

106. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0232076

107. Zhang A, Hünerbein M, Dai Y, Schlag PM, Beller S. Construct validity testing of a laparoscopic surgery simulator (Lap Mentor®): Evaluation of surgical skill with a

virtual laparoscopic training simulator. Surg Endosc Other Interv Tech. 2008;22(6):1440–4.

108. Matsuda T, McDougall EM, Ono Y, Hattori R, Baba S, Iwamura M, et al. Positive Correlation Between Motion Analysis Data on the LapMentor Virtual Reality Laparoscopic Surgical Simulator and the Results from Videotape Assessment of Real Laparoscopic Surgeries. J Endourol. 2012;26(11):1506–11.

109. McDougall EM, Corica FA, Boker JR, Sala LG, Stoliar G, Borin JF, et al. Construct Validity Testing of a Laparoscopic Surgical Simulator. J Am Coll Surg. 2006;202(5):779–87.

110. Woodrum DT, Andreatta PB, Yellamanchilli RK, Feryus L, Gauger PG, Minter RM. Construct validity of the LapSim laparoscopic surgical simulator. Am J Surg. 2006;191(1):28–32.

111. Schreuder HWR, van Dongen KW, Roeleveld SJ, Schijven MP, Broeders IAMJ. Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. Am J Obstet Gynecol. 2009;200(5):540.e1–8.

112. Brunner WC, Korndorffer JR, Sierra R, Massarweh NN, Dunne JB, Yau CL, et al. Laparoscopic virtual reality training: Are 30 repetitions enough? J Surg Res. 2004;122(2):150–6.

113. Gauger PG, Hauge LS, Andreatta PB, Hamstra SJ, Hillard ML, Arble EP, et al. Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. Am J Surg. 2010;199(1):72–80.

114. Larsen CR, Oestergaard J, Ottesen BS, Soerensen JL. The efficacy of virtual reality simulation training in laparoscopy: A systematic review of randomized trials. Acta Obstet Gynecol Scand. 2012;91(9):1015–28.

115. Pink DH. Drive: The Surprising Truth about What Motivates Us. Canongate Books Ltd. 2009.

116. Burden C, Oestergaard J, Larsen CR. Integration of laparoscopic virtual-reality simulation into gynaecology training. BJOG. 2011;118 Suppl 3:5–10

117. Stefanidis D, Heniford BT. The formula for a successful laparoscopic skills curriculum.

Arch Surg. 2009;144:77–82.

118. Ericsson KA. Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains. Acad Med. 2004;79(Suppl):S70–81.

119. Tsuda S, Scott D, Doyle J, Jones DB. Surgical Skills Training and Simulation. Curr Probl Surg. 2009;46(4):271–370.

120. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg. 2007;193(6):797–804.

121. Buks E, Schuster R, Heiblum M, Mahalu D, Umansky V. Dephasing in electron interference by a 'which-path' detector. Nature. 1998;391:871–4.

123. Larsen DP, London DA, Emke AR. Using reflection to influence practice: student perceptions of daily reflection in clinical education. Perspect Med Educ. 2016;5(5). 123. Nayar SK, Musto L, Fernandes R, Bharathan R. Validation of a virtual reality laparoscopic appendicectomy simulator: a novel process using cognitive task analysis. Ir J Med Sci. 2019 Aug 1;188(3):963–71.

124. Smink DS, Peyre SE, Soybel DI, Tavakkolizadeh A, Vernon AH, Anastakis DJ. Utilization of a cognitive task analysis for laparoscopic appendectomy to identify differentiated intraoperative teaching objectives. Am J Surg. 2012;203(4):540–5.

125. Bjerrum F, Strandbygaard J, Rosthøj S, Grantcharov T, Ottesen B, Sorensen JL. Evaluation of procedural simulation as a training and assessment tool in general surgery—simulating a laparoscopic appendectomy. J Surg Educ. 2017;74(2):243–50.

126. Brown C, Robinson D, Egan R, Hopkins L, Abdelrahman T, Powell A, et al. Prospective Cohort Study of Haptic Virtual Reality Laparoscopic Appendicectomy Learning Curve Trajectory. J Laparoendosc Adv Surg Tech. 2019;29(9):1128–34.

127. Beyer-Berjot L, Patel V, Acharya A, Taylor D, Bonrath E, Grantcharov T, et al. Surgical training: design of a virtual care pathway approach. Surgery. 2014;156(3):689–97.

128. Skjold-Odegaard B, Ersdal HL, Assmus J, Nedrebo BSO, Sjo O, Soreide K. Development and clinical implementation of a structured, simulation-based training programme in

laparoscopic appendectomy: description, validation and evaluation. BMJ Simul Technol Enhanc Learn. 2021;7(6):517–23.129.

129. Jeronimus BF, Laceulle OM. Frustration. Encyclopedia of Personality and Individual Differences. Cham: Springer International Publishing; 2017. 1–8 p.

130. Lazarus R, Folkman S. Stress, appraisal, and coping. New York: Springer; 1984.

131. Kaafarani HMA, Itani KMF, Giobbie-Hurder A, Gleysteen JJ, McCarthy M, Gibbs J, et al. Does Surgeon Frustration and Satisfaction with the Operation Predict Outcomes of Open or Laparoscopic Inguinal Hernia Repair? J Am Coll Surg. 2005 May 1;200(5):677–83.

132. Lieberman J, Dorey F, Shekelle P, Schumacher L, Thomas B, Kilgus D, et al. Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. J Bone Joint Surg Am. 1996;78(6):835–8.

133. Gostlow H, Marlow N, Babidge W, Maddern G. Systematic Review of Voluntary Participation in Simulation-Based Laparoscopic Skills Training: Motivators and Barriers for Surgical Trainee Attendance. J Surg Educ. 2017 Mar 1;74(2):306–18.

134. Burden C, Appleyard TL, Angouri J, Draycott TJ, McDermott L, Fox R. Implementation of laparoscopic virtual-reality simulation training in gynaecology: a mixed-methods design. Eur J Obstet Gynecol Reprod Biol. 2013 Oct 1;170(2):474–9.

135. van Dongen KW, van der Wal WA, Rinkes IHMB, Schijven MP, Broeders IAMJ. Virtual reality training for endoscopic surgery: voluntary or obligatory? Surg Endosc. 2008 Aug 18;22:664–7.

136. Blackhall VI, Cleland J, Wilson P, Moug SJ, Walker KG. Barriers and facilitators to deliberate practice using take-home laparoscopic simulators. Surg Endosc. 2019 Nov 19;33:2951–9.

137. Shetty S, Zevin B, Grantcharov TP, Roberts KE, Duffy AJ. Perceptions, Training Experiences, and Preferences of Surgical Residents Toward Laparoscopic Simulation Training: A Resident Survey. J Surg Educ. 2014 Sep 1;71(5):727–33.

138. Petrucci AM, Kaneva P, Lebedeva E, Feldman LS, Fried GM, Vassiliou MC. You Have a Message! Social Networking as a Motivator for FLS Training. J Surg Educ. 2015 May

1;72(3):542–8.

139. Van Empel PJ, Verdam MGE, Strypet M, Van Rijssen LB, Huirne JA, Scheele F, et al. Voluntary Autonomous Simulator Based Training in Minimally Invasive Surgery, Residents' Compliance and Reflection. J Surg Educ. 2012 Jul 1;69(4):564–70.

140. Chang L, Petros J, Hess DT, Rotondi C, Babineau TJ. Integrating simulation into a surgical residency program. Surg Endosc. 2007 Dec 16;21:418–21.

141. Stefanidis D, Acker CE, Greene FL. Performance Goals on Simulators Boost Resident Motivation and Skills Laboratory Attendance. J Surg Educ. 2010 Mar 1;67(2):66–70.

142. Korndorffer JR, Bellows CF, Tekian A, Harris IB, Downing SM. Effective home laparoscopic simulation training: A preliminary evaluation of an improved training paradigm. Am J Surg. 2012 Jan 1;203(1):1–7.

143. Csikszentmihalyi M. Beyond Boredom and Anxiety: Experiencing Flow in Work and Play. San Franisco: Jossey-Bass; 2000.

144. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. Neuron. 2010 Dec 9;68(5):815–34.

145. Aho JM, Ruparel RK, Graham E, Zendejas-Mummert B, Heller SF, Farley DR, et al. Mentor-Guided Self-Directed Learning Affects Resident Practice. J Surg Educ. 2015 Jul 1;72(4):674–9.

146. Stefanidis D, Acker CE, Swiderski D, Heniford BT, Greene FL. Challenges During the Implementation of a Laparoscopic Skills Curriculum in a Busy General Surgery Residency Program. J Surg Educ. 2008 Jan 1;65(1):4–7.

147. Verdaasdonk E, Dankelman J, Schijven M, Lange J, Wentink M, Stassen L. Serious gaming and voluntary laparoscopic skills training: A multicenter study. Minim Invasive Ther. 2009;18(4):232–8.

148. Seymour NE. Integrating simulation into a busy residency program. Minim Invasive Ther Allied Technol. 2005;14(4–5):280–6.

149. Kim S, Ross B, Wright A, Wu M, Benedetti T, Leland F, et al. Halting the revolving door of faculty turnover recruiting and retaining clinician educators in an academic medical simulation center. Simul Healthc. 2011 Jun;6(3):168–75.

150. Hanks JB. Simulation in Surgical Education: Influences of and Opportunities for the Southern Surgical Association. J Am Coll Surg. 2019;228(4):317–28.

151. Kordowicz AGR, Gough MJ. The challenges of implementing a simulation-based surgical training curriculum. Br J Surg. 2014 Apr;101(5):441–3.

152. Jamal MH, Rousseau MC, Hanna WC, Doi SAR, Meterissian S, Snell L. Effect of the ACGME duty hours restrictions on surgical residents and faculty: A systematic review. Acad Med. 2011;86(1):34–42.

153. Waljee JF, Chung KC. Academic Plastic Surgery: Faculty Recruitment andRetention. Plast Reconstr Surg. 2014 Mar;133(3):407e.

154. Arora J, Makker R, Kerr B. P21 A simulation faculty recruitment crisis: are simulation fellows the answer? BMJ Simul Technol Enhanc Learn. 2018;4(Suppl 2):A60–1.

155. Loughrey M, Naidoo U. P15 Is there a problem finding and retaining clinical faculty for foundation doctor simulation in south london, kent, surrey and sussex? BMJ Simul Technol Enhanc Learn. 2018 Nov 1;4(Suppl 2):A58 LP-A58.

156. Sharma M, Horgan A. Comparison of Fresh-Frozen Cadaver and High-Fidelity Virtual Reality Simulator as Methods of Laparoscopic Training. World J Surg. 2012;36:1732–7.

157. Association for Simulated Practice in Healthcare. Simulation-based education in healthcare. Standards framework and guidance. 2016.

158. The Royal College of Surgeons of England. Improving Surgical Training. RCS Professional Standards; 2015.

160. Powell C. The Delphi technique: myths and realities. J Adv Nurs. 2003;41(4):376–82.

161. Bongers PJ, Van Hove PD, Stassen LPS, Dankelman J, Schreuder HWR. A new virtual-reality training module for laparoscopic surgical skills and equipment handling: Can multitasking be trained? A randomized controlled trial. J Surg Educ. 2015;72(2):184–91.

List of Abbreviations

2D	2-Dimensional
3D	3-Dimensional
ACS	American College of Surgeons
AERA	American Educational Research Association
APA	American Psychological Association
ASPiH	Association for Simulated Practice in Healthcare
BS	Basic Skills
CCT	Certificate of Completion of Training
CO_2	Carbon dioxide
CSBS	Centre for Screen-Based Simulation
CT1	Core Trainee (year 1)
CT2	Core Trainee (year 2)
CTA	Cognitive Task Analysis
DP	Deliberate Practice
EWTD	European Working Time Directive
FLS	Fundamentals of Laparoscopic Surgery
FP	Full Procedure
FY	Foundation Year
GBP	United Kingdom Pound Sterling
GI	Gastrointestinal
GMC	General Medical Council
GOALS	Global Operative Assessment of Laparoscopic Skills
HEE	Health Education England
ISCP	Intercollegiate Surgical Curriculum Programme
JCST	Joint Committee on Surgical Training
LA	Laparoscopic Appendicectomy
MMC	Modernising Medical Careers
NCEPOD	National Confidential Enquiry into Patient Outcomes and Deaths
NCME	National Council of Measurement in Education
OR	Operating Room
OSATS	Objective Structured Assessment of Technical Skill
PGY	Post-Graduate Year
PT1	Procedural Task 1
PT2	Procedural Task 2
PT3	Procedural Task 3
PT4	Procedural Task 4
PT5	Procedural Task 5
SAGES	Society of American Gastrointestinal Endoscopic Surgery

SBME	Simulation-Based Medical Education
SHO	Senior House Officer
SpR	Specialist Registrar
ST	Specialist Trainee
StR	Specialty Registrar
TEL	Technology-Enhanced Learning
UCLMS	University College London Medical School
UGRA	Ultrasound-Guided Regional Anaesthesia
UK	United Kingdom
USD	United States Dollar
VR	Virtual Reality
WHO	World Health Organization

Appendix i

Induction Check-list

Stage 1 (15 minutes):

Pre-study introduction, questionnaire and consent form

1. Participant to read ☐
2. Faculty to follow through with the same information/instructions verbally ☐
3. Opportunity for participant to ask questions ☐
4. Participant to sign consent form ☐
5. Participant to fill in questionnaire ☐

The first and subsequent sessions

6. Start on a day separate to induction day ☐
7. How to book a session
    a. Call Royal Free Hospital, x36857 and/or p.berlingieri@ucl.ac.uk – liaise with Dr Pasquale Berlingieri. He must be present during practice sessions ☐
8. Sign the register upon the start of each simulator session ☐
9. Must complete the first two repetitions on the first day ☐

Safety of participants

10. What to do in the event of a fire ☐
11. Location of the nearest exits ☐

Stage 2 (20 minutes):

The equipment

12. Username and password, assign module ☐
13. How to switch the machines on and off safely ☐
14. How to log on to the machines ☐
15. Mobile phone use in the simulator room is not permitted and should be switched off ☐
16. No food or drink is allowed in the simulator room ☐
17. Each session must be uninterrupted with no talking ☐

Demonstration

18. Which tasks the participant must do, and in which order ☐
19. How to choose/select the tasks ☐
20. How to calibrate the instruments ☐
21. Use of roticulator ☐
22. Demonstration of each psychomotor task ☐
23. How to view the performance data ☐
24. What to do if there is a computer error / malfunction ☐
25. Final opportunity for participant questions ☐

Appendix ii

Study Introduction

Here at The Royal Free Medical Simulation Centre we are devising a novel laparoscopic appendicectomy curriculum for the acquisition of key laparoscopic skills using our state-of-the-art Lap Mentor™ Virtual Reality (VR) simulators. Your participation in this study helps us to devise a scientifically valid method for medical students and trainees to quickly become proficient at laparoscopy. Studies have already demonstrated significant improvement in laparoscopic skills following training on VR laparoscopic simulators, and so we expect you to benefit from taking part.

During this induction you will have been randomised to one of three groups. As you are in group A, you have been assigned to practice the psychomotor skills tasks on the Lap Mentor™.

In order for this to be successful for both you and our study, you **must** abide by the following rules:

1. Starting with psychomotor task 1 and moving sequentially, complete all 9 psychomotor skills tasks in one session;
2. No session is allowed to last longer than 45 minutes;
3. You must complete 10 repetitions of each task in total (thereby completing at least 10 sessions);
4. There must be a gap of at least one hour in between practice sessions;
5. You must not perform more than two practice sessions per day;
6. Your first two sessions must be during the same day;
7. You **must complete and save every attempt** at each task by clicking 'finish' upon task completion, no matter how good or bad you think it was. This means, for example, that if you feel you started off poorly, you must not 'start over', and you must continue to the best of your ability until task completion. This does not affect your success at course completion (the scores are anonymised). This is **absolutely critical**;
8. You must not practice with any other simulator until you finish the course;
9. You must not receive any tuition on the simulator from **anybody**;
10. You must not have any real laparoscopic operating experience during the study period. If any experience is offered, you must decline and you may cite our study protocol. The exemptions to this are (and please inform us):
    a. Where your input is necessary for the care of the patient
    b. Camera navigation (i.e. holding the camera)
11. You have 3 weeks to finish the course;
12. How to book a session:
    a. Call Royal Free Hospital, x36857 and/or p.berlingieri@ucl.ac.uk – liaise with Dr Pasquale Berlingieri. He must be present during practice sessions
13. Sign the register upon the start of each simulator session.

In return for following our rules, and for your successful completion of this course, we will provide you with a Certificate of Course Completion for Virtual Reality Laparoscopic Psychomotor Skills.

If you would like to go ahead then please fill in our pre-course questionnaire, and read/sign the declaration.

Appendix iii

Pre-Study Questionnaire

1. Name: ……………………………………………

2. Email address: …………………………………………………

3. Mobile telephone number: ………………………………………………

4. Postal address:

………
……………………………

5. Age: ……………

6. Sex:
   - ☐ M
   - ☐ F

7. Dominant hand:
   - ☐ Right
   - ☐ Left
   - ☐ Both hands have equal dominance

8. Current level of training:
   - ☐ FY1
   - ☐ FY2

9. Current specialty: ……………………………

10. Desired career path:
    - ☐ medicine
    - ☐ general practice
    - ☐ surgery
    - ☐ undecided
    - ☐ other, please specify: ……………………………

11. How certain are you of the above career selection? (1 = "just an idea", 5 = "100% certain"):

    ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

12. No. of courses attended in last 12 months (please specify): ………

13. Which exam have you taken or are currently studying for?

    ☐ None/not studying currently ☐ MRCP ☐ MRCS

☐Other:…………………

14. How long would it take for you to get to the Royal Free Hospital from **home**?
☐ 0-15m ☐ 16-30m ☐ 31-45m ☐ 46m-1hr ☐ >1hr

15. How long would it take for you to get to the Royal Free Hospital from **work**?
☐ 0-15m ☐ 16-30m ☐ 31-45m ☐ 46m-1hr ☐ >1hr

16. How much time, in total, have you previously spent practicing on a mechanical laparoscopic box trainer (or laparoscopic 'video trainer')?

………

17. How much time, in total, have you previously spent practicing on the Simbionix Lap Mentor™ virtual reality laparoscopic simulator (computer-generated graphics, with force feedback)?

………

18. How much time, in total, have you previously spent practicing on any **other** virtual reality laparoscopic simulator?

………

19. How many full laparoscopic procedures (excluding diagnostic laparoscopy) have you performed (as *primary operator*)?

………

20. How many full **laparoscopic appendicectomies** have you performed (as *primary operator*)?

………

21. How many times have you been the **camera navigator** for **any** laparoscopic procedure?

………

22a. Which musical instrument do you play (if any)?	22b. How many hours per week do you **currently** play this instrument?	22c. How long have you been playing this instrument regularly?

23. How many hours per week do you **currently** play computer games?
☐ 0 ☐ 0-2 ☐ 2-4 ☐ 4-7 ☐ >7

24. What is your **past** experience of playing computer games?
☐ never
☐ rarely played
☐ used to occasionally
☐ used to regularly

25. Which sport do you play (if any) more than once per month? ……………………

Appendix iv

Participant Consent & Declaration*

By signing below, you are consenting to the use of your data (on this questionnaire and from the stored performance data on the laparoscopy simulators) for the purpose of research, which may be published and/or presented. No personalised details will be used in the case of publication or presentation.

By signing, you are also declaring that you will:

a) Complete and save every performance on the simulator, no matter how good or bad (this will not affect your outcome);

b) refrain from other simulated laparoscopic experiences until you have completed the course;

c) decline any real laparoscopic operating experience, other than camera navigation or where patient care depends upon it**;

d) Abide by the rules outlined in the 'Study Introduction'.

Signed:

Print name:

Date:

** This form was used for Part 1 of the study.*

*** This clause was present only on novice participants' consent & declaration form.*

Appendix v

Participant Consent & Declaration*

By signing below, you are consenting to the use of your data (on this questionnaire, on the paper based diary and from the stored performance data on the laparoscopy simulators) for the purpose of research, which may be published and/or presented. No personalised details will be used in the case of publication and/or presentation.

By signing, you are also declaring that you will:

a) complete and save every performance on the simulator, no matter how good or bad;

b) refrain from other simulated laparoscopic experiences until you have completed the course;

c) decline any real laparoscopic operating experience, other than camera navigation or where patient care depends upon it;

d) abide by the rules outlined in the ‘Study Introduction’.

Signed:

Print name:

Date:

** This form was used for Part 3 of the study.*