1-1-2014

The impact of innate ability and objective scoring methods on proficiency based training in minimally invasive surgery.

Christina Buckley
Royal College of Surgeons in Ireland

The Impact of Innate Ability and Objective Scoring Methods on Proficiency Based Training in Minimally Invasive Surgery

A thesis submitted to the Royal College of Surgeons in Ireland, in fulfillment of the requirements for the degree of Doctor of Medicine (MD)

RCSI

by

Dr. Christina Buckley MB BCh BAO (Hons.), MRCS

Supervised by

Professor Paul Neary, MD, FRCSI

Vice Dean of Dublin Region, Postgraduate Surgical Training, RCSI

&

Professor Oscar Traynor, MCh, FRCSI

Dean of Postgraduate Surgical Training, RCSI

List of Figures

List of Tables

List of Abbreviations

MIS Minimally Invasive Surgery

EWTD European working time directive

NSTC National surgical training centre

RCSI Royal College of Surgeons in Ireland

VR Virtual reality

PBPT Proficiency based progression training

OSATS Objective structured assessment of technical skill

BST Basic surgical training

HST Higher surgical training

BST's Basic surgical trainees

HST's Higher surgical trainees

MCQ Multiple-choice questions

PicSOr Pictorial surface orientation test

CELTS Computer enhanced laparoscopic training system

St. Dev. Standard Deviation

ICC Intraclass correlation coefficient

Abstract

Background: The attainment of technical competence and accurate performance assessment of surgical trainees for surgical procedures are the fundamental components of a proficiency-based surgical training programme. We hypothesised that aptitude may directly affect one's ability to successfully complete the learning curve for minimally invasive procedures.

Aim: The principle aim of this thesis was evaluate the impact of innate ability upon the rate at which a surgical novice can achieve proficiency in index and advanced laparoscopic procedures. Our secondary aim was to develop new objective methods of technical skills assessment for a proficiency-based programme.

Materials & Methods: We tested medical students (surgical novices) with disparate aptitude consecutively until they achieved proficiency in laparoscopic appendicectomy and laparoscopic suturing using objective and subjective scoring methods. We developed objective scoring methods by designing a new zone metric to assess laparoscopic suturing and also a mathematical formula to provide meaningful metrics scores on the laparoscopic simulator ProMIS.

Results: The results demonstrated that surgical novices with low aptitude took twice as long to reach proficiency targets. Aptitude predicted superior baseline performance in medical students. There is a group of surgical candidates who are unable to achieve proficiency despite repeated practice. It was shown that a new zone metric could be used to assess laparoscopic suturing. Finally we successfully developed a scoring method, which provides meaningful user scores on the ProMIS simulator.

Conclusion: High aptitude is directly related to a rapid attainment of proficiency. It is likely that surgical trainees self select in surgery based on innate ability. The new zone metric and formulated scoring systems are valid tools for assessing laparoscopic tasks and provide meaningful scores. These findings have implications for developing a proficiency based training system according to a trainees natural ability.

Acknowledgements

Supervisors

Professor Paul Neary

Professor Oscar Traynor

Additional Mentors

Mr. Dara Kavanagh

Mr. Sean Carroll

Professor Frank Keane

Mr. Tom Gallagher

Professor Sean Tierney

Statistician

Dr Ronan Conroy

In Acknowlegement of their Generous Support

Mr. Donncha Ryan

Ms. Roisin Dolan

Mr. Dave Matthews

Ms. Emmeline Nugent

Ms. Siobhan O'Keeffe

Mr. Kieran Taugney

In Acknowledge of their time and effort

All the Medical Students, BST's, HST's and Consultant Surgeons who participated

Special Thank You

Dara Kavanagh and Donncha Ryan for their invaluable support and dedication to the completion of this work

Special Acknowledgement

Christopher, Victoria, Andre and Laura Buckley

Dedicated to

Bernard Eugene Buckley

Peer Reviewed Publications

Buckley CE, Kavanagh DO, Nugent E, Ryan D, Traynor OJ, Neary PC.

Zone calculation as a tool for assessing performance outcome in laparoscopic suturing.

Surg Endosc. Oct 2014 PMID: 25303906

Buckley CE Nugent E, Neary PC, Traynor OJ, Carroll SM.

Do plastic surgical trainees naturally self-select based on fundamental ability?

JPRAS May 2014 PMID: 24880578

Buckley CE, Kavanagh DO, Nugent E, Ryan D, Traynor OJ, Neary PC.

The impact of aptitude on the learning curve for laparoscopic suturing.

Am J Surg. Feb 2014. PMID: 24468026.

Buckley CE, Kavanagh DO, Traynor OJ, Neary PC.

Is the skillset obtained in surgical simulation transferable to the operating theatre?

Am J Surg. Jan 2014 PMID: 24238602.

Buckley CE, Kavanagh DO, Gallagher TK, Conroy RM, Traynor OJ, Neary PC.

Does aptitude influence the rate at which proficiency is achieved for laparoscopic appendectomy?

J Am Coll Surg. Dec 2013. PMID: 24051067

List of Publications and Communications

Book Chapter

Buckley CE, Nugent E, Ryan D, Neary PC

Virtual Reality – A New Era in Surgical Training,

In: *"Virtual Reality in Psychological, Medical and Pedagogical Applications"*

ISBN 978-953-51-0732-3. Published September 12, 2012.

Oral Presentations

"Assessment of Virtual Reality Zone Training as a Novel Tool in Laparoscopic Suturing Evaluation"

CE Buckley, E Nugent, D Ryan, O Traynor, PC Neary

Presented at the "37th Sir Peter Freyer Surgical Symposium, Galway, September 2012".

"Does Aptitude Matter?"

CE Buckley, DO Kavanagh, R Conroy, O Traynor, PC Neary

Presented at the "American College of Surgeons 98th Annual Clinical Congress (ACS), Chicago, Illinois, October 2012'.

"Do Plastic Surgical Trainees Naturally Self-Select Based on Fundamental Ability?"

CE Buckley, E Nugent, PC Neary, O Traynor, SM Carroll

Presented at the "Irish Association of Plastic Surgery Winter Meeting, November 2012"

"The Impact of Aptitude on the Learning Curve for Laparoscopic Suturing"

CE Buckley, DO Kavanagh, E Nugent, D Ryan, O Traynor, PC Neary

Presented at the "Association for Surgical Education (ASE), Orlando, Florida. April 2013".

Poster Presentations

"Does Innate Ability Dictate the Learning Curve for Laparoscopic Suturing"

CE Buckley, E Nugent, PC Neary, O Traynor

Presented at the "37th Sir Peter Freyer Surgical Symposium, Galway, September 2012".

"Innate Ability of Surgical Trainees: Does the Cream always come to the Top?"

E Nugent, **CE Buckley**, O Traynor, PC Neary

Presented at the "37th Sir Peter Freyer Surgical Symposium, Galway, September 2012".

I declare that this thesis, which I submit to RCSI for examination in consideration of the award of a higher degree (MD) is my own personal effort. Where any of the content presented is the result of input or data from a related collaborative research programme this is duly acknowledged in the text such that it is possible to ascertain how much of the work is my own. I have not already obtained a degree in RCSI or elsewhere on the basis of this work. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

Signed ______________________

Student Number 11112671

Date 21.10.14

Chapter 1

Introduction

1.1 New Era of Laparoscopy

Since the adoption of laparoscopy in the 1990's, surgical techniques have been revolutionised (Satava, 1999). The advent of laparoscopy and its basis for minimally invasive procedures has provided many benefits for the patient. The ability to access the abdomen in a safe, quick, simple and minimally invasive fashion is very advantageous. Reduced post-operative pain, better cosmesis, quicker recovery post-operative time and reduced length of stay are all well documented benefits of the laparoscopic approach (Gurusamy et al., 2008, Sauerland et al., 2010). Furthermore, benefits to our healthcare system as a whole have become apparent which include lower morbidity, greater number of day cases being performed, shorter period of in-hospital post-operative care (Delaney et al., 2008, Kehlet, 2008) culminating in an increased utilisation of in-patient beds. Laparoscopy has evolved during a period characterized by a radical evolution of patient expectations, appropriate emphasis on patient safety and a call for competency-based training rather than the traditional apprentice-based model.

1.2 Current Challenges in Surgical Training

1.2.1 Acquiring a New Skill Set

The skill set required to perform laparoscopy is very different compared with open surgery (Figert et al., 2001). Laparoscopy is associated with a lack of tactile feedback, greater emphasis on precise hand eye coordination, and a change from 3-dimensional (3-D) to 2-dimensional (2-D) visualisation as well as adaption to the fulcrum effect (Gallagher et al., 1998, Perkins et al., 2002). The skill set cannot be taught easily in the real life environment under the supervision of a senior surgeon. With the traditional open approach the supervising surgeon can directly guide the hands of the trainee and immediately intervene if a problem or difficulty arises. In laparoscopy however the expert surgeon has less control over what the trainee is doing. If a complication were to arise during the course of the surgery it would be more difficult for the expert surgeon to intervene and rectify the situation. The same can be said for endovascular procedures and endoscopy. The learning curve is also steeper in minimally invasive surgery (MIS) than for open surgery (Rosser et al., 1997, Rogers et al., 2001).

1.2.2 Steep Learning Curve

The early part of the laparoscopic learning curve is associated with a higher complication rate (Buckley et al., 2014b). Therefore it is intuitive that familiarity with surgical procedures should be taught outside the surgical environment in order to improve patient's safety. Laparoscopic cholecystectomy was the index procedure for laparoscopy. Although it was embraced with vigor it was also the procedure where problems and concerns with the minimally invasive approach were first highlighted. A higher than acceptable rate of bile duct injury in laparoscopic cholecystectomy when compared to open cholecystectomy became an important issue in the 1990's.

The Southern Surgeons Club study is an oft referenced paper (Moore and Bennett, 1995) . They found that 90% of common bile duct injuries occurred within the first 30 operations performed by the trainee surgeon. They also predicted that the surgeon had a 1.7% probability of causing a bile duct injury in their first operation, which reduced to 0.17% by the 50th case. The probability of injury was found to have dropped to a significantly safe level by the 10th case (Moore and Bennett, 1995, Wherry et al., 1996). This was one of the first articles to underline the significance of the learning curve in minimally invasive surgery.

As the complexity of the procedure increases so too does the learning curve. This has been demonstrated for laparoscopic fundoplication, where a significant reduction in complications has been reported to reduce only after the 50th case with the highest complication rate found within the first 20 cases (Watson et al., 1996). The learning curve for laparoscopic colectomy has been estimated to be even higher, with the highest rate of complications occurring during the first forty procedures (Bennett et al., 1997). The initial learning curve has been shown to be associated with the period of greatest risk to the patient.

1.2.3 Patient Safety and Delivery of Healthcare

Advancement of minimally invasive surgery is not the only challenge to surgical training in Ireland. A global awareness of patient safety has evolved over the past decade (I., 2001, Kohn LT, 1999). Attitudes of patients, doctors and the population as a whole have changed. There has been a major change in recent years in the structure of our healthcare system. The introduction of the European Working Time Directive (EWTD) has reduced the training opportunities for trainee surgeons. This has a direct impact upon development of surgical skills (Crofts et al., 1997, Skidmore, 1997). The model advocated by recent government places emphasis upon consultant-delivered healthcare. However recent healthcare strategies have placed emphasis on service provision. The emphasis of recent health systems on this consultant-led concept has modified public perception such that patients feel entitled to undergo surgery under the care of their named consultant. This provides a major challenge to the traditional apprentice teaching module.

Economic factors also affect training structures. Time pressures in the operating theatre have hindered the amount of operative experience a trainee receives. This is due to the global economic recession as financial restraints on supervising surgeons are greater than before. Healthcare systems reward efficiency in use of the operative room, which creates a dynamic high-turnover model not conducive to development, and nurturing of technical skills (Babineau et al., 2004, Bridges and Diamond, 1999).

1.2.4 Current Training Environment

The current training environment is a very different one to the traditional approach of "see one, do one, teach one" which William Halstead introduced in 1889 in John Hopkins Hospital (Kerr and O'Leary, 1999). This traditional apprenticeship model of training no longer applies in an era of minimally invasive surgery and evolving patient expectations. Trainees are reliant on the enthusiam of their supervising surgeon, the case load and variation as well as the location of the hospital in order to receive adequate exposure to a variety of surgical techniques and procedures.

Further to this, the current training paradigm lacks objective feedback on trainee performance. Training needs to be done in a more efficient manner to optimize the learning experience and surgical exposure of the trainee as well as combating the challenges we face in adapting the new skill set required with minimally invasive procedures (Wolfe et al., 1993). For these reasons, the surgical community looked to virtual reality as a way of bridging this skill gap and providing a method of safely introducing new techniques into surgical practice.

1.3 How Simulation-Based Training offers a Solution

Training bodies have aimed to address these aforementioned issues through development of simulated education programmes. One of the first steps towards simulated surgical training in this country was the development of a state of the art simulation laboratory in the National Surgical Training Centre (NSTC) in the

Royal College of Surgeons in Ireland (RCSI) in 2003. They also developed the first international mobile skills unit to provide teaching opportunities at remote sites throughout the country. Previous work done in our institution (Boyle et al., 2011, Neary et al., 2008, Nugent et al., 2013) have demonstrated the usefulness of simulators in surgical training.

Satava was the first to recommend simulation as a complement to current training models (Satava, 1993). Using simulation to improve surgical skill is now a very acceptable method of training (Aggarwal et al., 2006b, Ahlberg et al., 2007, Fried et al., 2004, Korndorffer et al., 2005). The introduction and development of virtual reality (VR) and hybrid simulators has been one of the main innovations that have resulted in a change in the surgical training curriculum. A simulation laboratory is a space designated for trainees to practice various skills and procedures on a wide variety of available surgical simulators in a safe, controlled environment. Dedicated time is necessary in order to learn the required skills in a protected manner. Simulation has much more to offer the trainee than the clinical environment alone as it allows for dedicated teaching which is focused and structured with specific learning goals. Therefore it is appropriate that we institute carefully thought out, well-structured curricula. Several studies (Aggarwal et al., 2009, Gallagher et al., 2005, Kolozsvari et al., 2011a) have proposed templates for this.

The role of simulation in surgery is to provide our trainees with the opportunity to learn basic tasks in a safe and controlled environment. All movements the trainee makes can be recorded and thereby facilitating immediate and objective feedback. It is also possible to set a proficiency level on a simulator and therefore design a training program giving set goals that a trainee needs to accomplish before being allowed perform in the operating theatre. All of these factors contribute to skill learning, assessment, selection and credentialing. The use of simulation should provide the setting in which challenges such as the use of new instruments and technology can be overcome. An example of this is single incision laparoscopic surgery where the 'triangulation' approach, which is intrinsic to minimally invasive surgery, is absent. Partial compensation is achieved by the development of curved instruments, which generate counterintuitive movement, which can reproduce a form of triangulation.

Given that simulation is generally an education tool, there are two distinct parts to the delivery of training on simulators. There is firstly the teaching aspect which is the way which we communicate or impart knowledge or information. Secondly there is the training aspect, which is the acquisition and development of psychomotor skill and cognitive skill (Gallagher et al., 2005). By mastering skills such as hand-eye coordination, counter intuitive fine movements and the ability to work with a 2-D image in a 3-D space on a simulator, the trainee surgeon can then focus on the critical steps of the operation when in the operating theatre. One of the difficulties with acquiring these skills is due to the fulcrum effect of the body wall on instrumentation (Gallagher et al., 1998). This problem cannot be overcome with concentration. It requires practice until the process becomes automated (Gallagher and Satava, 2002). This process of fragmenting the education process replaces the conventional learning model where all components are assimilated simultaneously.

1.4 Role of Simulation in Surgical Training

1.4.1 Historical Background

Simulation has its roots in the commercial and military aviation industry. It was first considered in 1910 when student pilots trained in a land-borne aircraft with reduced wingspans. The first rudimentary simulator was available in 1929 and was known as the Links Trainer (LL, 1970). It consisted of a wooden fuselage mounted on an air bellows, which was able to represent the movements involved in flight. This allowed the pilot to train for hours. In 1934 the US purchased six Links simulators following a series of aviation accidents. At this stage it was recognized that the current training programs were inadequate and simulation was a step towards improving the training system. World War II also had a dramatic impact on the uptake of simulation for training purposes. The war demanded that a greater number of pilots be trained and that skills such as the need to become proficient in instrument or blind flying were paramount. These factors led to simulator development and usage. Today, they have hugely sophisticated systems which replicate an aircraft's environment precisely and can deal with a vast range of potential flight scenarios. Pilots must undergo ongoing training annually entitled "checking out" by the Federal Aviation Administration in order to ensure ongoing certification as well as additional training requirements if they wish to change to another aircraft. Astronauts are also required to follow the same procedures.

The first surgical simulator to use virtual reality technology was created at NASA by Rosen and Delp (Delp et al., 1990). It was an orthopaedic lower limb model that simulated tendon transfer. It was unique in that it allowed planning and therefore optimisation of operations. Virtual reality technology has evolved to the point today where patient data and radiological images can be in-putted into the simulator allowing for a complete simulated run-through before operating on the patient, this process is known as 'mission rehearsal'.

The aviation industry paved the way for simulated surgical training. However simulating the human body is complex and extremely challenging due to our anatomical variance and the unpredictable nature of our physiology. Because of this, we are only able to reproduce certain aspects of a surgical procedure rather then replicating it entirely. Currently available simulators are suitable for reliable repetition of conditions and interface. The purpose is generally to allow the user to practice their skill in a controlled environment, with the additional benefit of having 'metrics' or computerised feedback on their performance.

1.4.2 Surgical Simulators

1.4.2.1 VR Trainers

VR Trainers digitally recreate the procedures and environment of laparoscopy. Jaron Lanier a philosopher and scientist coined the term “virtual reality” in the 1980’s. It is a phrase used to describe the concept of a virtual world, which supports interaction instead of something that is passively visualised. There is a wide range a commercially available VR surgical trainer’s including the LapSim, MIST-VR and LapVR. Each measures a variety of metrics and has differing levels of difficulty as well as varying complexity in graphics. Some examples of metrics measured include tool to tool contact, loss of tissue-tool contact, inappropriate “passing of the point” of the instrument through the tissue, inappropriate target release, inappropriate cautery application, economy of movement, time and procedure specific errors.

Virtual reality is an acceptable way of simulating a surgical procedure however there are several challenges given the limitations of modern technology. Graphics can simulate anatomical structures visually however they are unable to model the physical properties of human tissues. The lack of haptic feedaback remains a significant technological challenge in VR simulation. Currently none of the VR simulators are capable of providing any tactile feedback. There is ongoing research into this area however haptic technology is currently very rudimentary.

1.4.2.2 Hybrid Trainers

Because of the limitations of VR Trainers to provide tactile feedback, hybrid trainers were developed which combine computerized interfaces with ex vivo synthetic parts to provide tactile feedback. Haptics as a result does not pose a problem as you have a realistic simulated environment. The limitations of such physical models however include the increase in cost as the models can only be used once. Also complex human anatomy and physiology cannot be replicated precisely, for example bleeding vessels and leaking structures following trauma, and appropriate surrounding anatomy. Virtual reality surpasses physical models in this realm.

1.5 Methods of Assessment

Assessing improvement in a candidates surgical skill objectively is essential for monitoring progression through the surgical training pathway.

1.5.1 Metrics

Time is the most basic metric which may indicate progression in a task however it is not a true indicator of progression (Botden et al., 2009a, Emam et al., 2000). When time is combined with an error score (the amount of errors committed per task by the user), a trainee can also be assessed for accuracy. Early box trainers lacked tracking systems which recorded errors and time; a simple stopwatch was used to access the speed at which a task was performed. With the advent of virtual reality simulators, we now have stand alone systems which can measure and record metrics. Simulators can generate a profile summary upon completion of a procedure or task which provides immediate feedback and an opportunity to see ones progress upon repeated practice. The easy to use nature of VR simulators along with practice sessions and step by step instructions provides the user with an opportunity for practice and attainment of proficiency.

Further to basic metrics (time, errors), more sophisticated markers of performance measurement have emerged over the years. An example of this is instrument path length which is the distance travelled by the instrument or the sum of deviations from a fixed point. When this is applied to laparoscopy, this suggests operative focus and greater overall performance and experience. A study by Smith et al used computer sensors on the tips of laparoscopic instruments to track motion paths (Smith et al., 2001). They found that speed did not equate to improved performance hence time can be a misleading if not used in conjunction with other metrics. Another metric used is economy of movement which is a score based on sudden changes in acceleration that works as an indication of smooth movement or instrument handling.

In order to use metrics produced by simulators as assessment tools, they must undergo appropriate validation. There are three different types of validity. Construct validity is the ability of a simulator to detect differences between groups with different levels of experience. It order to validate any new metric, construct validity must be established (i.e the simulator can measure what it claims to measure). When introducing a new method of assessment/training in a simulated setting, the new method must be compared to a known standard. This is called concurrent validity and is defined as the concordance of a test to a known "gold standard". Face validity is the extent to which simulation resembles the real task, this is important for the trainees using the simulators.

In order to use simple metrics to measure proficiency, appropriate scoring systems must be developed. The computer enhanced laparoscopic training system (CELTS) was developed by the Centre for the Integration of Medicine and Innovative Technology CIMIT and Harvard Medical School (Stylopoulos et al., 2003). They used a box trainer with a computer interface to form a task-independent scoring system against expert benchmark levels. Expert scores were calculated for suturing, peg transfer and knot tying using time, path length, smoothness, and depth perception as metrics. The user's score was then compared with an expert score which led to the development of a standardised scoring system. This scoring method provided a gold standard of comparing novices to experts (Stylopoulos et al., 2004). When ProMIS was later developed, it contains system which can compare the user's score to expert proficiency scores on a bar chart for time, economy of movement and path length metrics. The scores need to be preset once they have been established by experts for each module. ProMIS is a hybrid simulator which enables the user to use physical models, which ensures appropriate tactile feedback. It tracks the instruments, thereby producing a report of metric results, which provides immediate feedback on performance.

1.5.2 Global Rating Scales

Further to metrics, subjective rating of a surgical performance remains a very important tool. An approach to testing operative skills outside the operative setting led to the Objective Structured Assessment of Technical Skill (OSATS) which was introduced by Reznick et al in 1996 (Faulkner et al., 1996). This seven-item table of technical performance on a five-point grading scale includes respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, knowledge of specific procedure and use of assistants (Appendix I). The OSATS tool has demonstrated high reliability and construct validity and is now used as a globally validated rating scale (Ault et al., 2001, Faulkner et al., 1996, Martin et al., 1997, Reznick et al., 1997).

Global assessments are now widely used in the assessment of proficiency during training and are used to study the effect that simulated surgical training has on operative skill. Several studies demonstrating the transfer of skill from a simulated environment to the operating room have used a slightly modified version of OSATS with an included parameter of overall performance (Hamilton et al., 2001, Lucas et al., 2008, Scott et al., 2000, Traxer et al., 2001).

A study by Grantcharov (Grantcharov et al., 2004) modified the scale (Appendix II) so that a new parameter was created - economy of movement, which was a combination of time and motion (1= clear economy of movements and maximum efficacy; 5= many unnecessary moves) and instrument handling (1= fluent moves with instruments; 5= repeated tentative awkward or inappropriate moves). In Reznick's original scale, five was the best possible score and one was the worst. In this study, a parameter of error score was also created which is a combination of respect for tissue from Renwick's scale (1=consistently handled tissues appropriately with minimal damage; 5= frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments) and precision of operative technique which is a new parameter (1= fluent, secure and correct technique in all stages of the operative procedure; 5= imprecise, wrong technique in approaching operative intentions). The Global Assessment of Laparoscopic Skills (GOALS) tool was designed by Vassiliou et al (Vassiliou et al., 2005) for minimally invasive procedures. This five-point scale assessed depth perception, bimanual dexterity, efficiency, tissue handling and autonomy. Results have shown that the tool is reliable and valid (Gumbs et al., 2007, Vaillancourt et al., 2011).

There is a trend towards using global rating tools in video analysis rather than direct observation in a live surgical setting due to time and cost resources. The advantage of simulation in this setting is the convenient storage of vast amounts of data. As there are so many available ways of rating surgical performance, the question of which is superior has been evaluated. A study by Aggarwal et al

(Aggarwal et al., 2007a) assessed four different scales, OSATS, modified OSATS with four instead of seven parameters, a procedure-specific global rating scale and a procedure checklist using laparoscopic cholecystectomy. The generic global rating scales successfully distinguished between novices and experts unlike the procedure specific rating scale or checklist. An extensive systematic review was undertaken by van Hove and colleagues to examine the current evidence for objective assessment methods for technical surgical skills (van Hove et al., 2010). It was concluded that OSATS is presently most accepted as the "gold standard" for objective skill assessment however it remains unknown whether OSATS can distinguish between different levels of performance. Furthermore cut off values have not been determined for OSATS. The same short comings apply to procedure specific checklists and currently there is only one checklist with a high level of evidence (Sarker et al., 2006). The study also concluded that motion analysis devices can determine between operators with different levels of experience. An important point that was discussed in this study is that the value of a good assessment method can diminish when it is used in an inappropriate setting.

1.6 Curriculum Design

1.6.1 Proficiency Based Training

The widespread implementation of the laparoscopic approach has prompted an explosion of literature recommending a specific number of cases for a given procedure to safely complete the learning curve and attain proficiency (Botden et al., 2009b, Simons et al., 1995, Tekkis et al., 2005). In fact they have not defined the additional volume required to attain the desired 'expert' level. The development of simulation-based training has attempted to address this. The development of proficiency in performing a specific procedure on a simulator prior to clinical performance is an ideal model of training which addresses the limiting training opportunities currently available to trainees. Allowing the surgical trainee to become proficient in surgical skills in a simulated lab is an attractive option. It teaches the basic skills of instrument and tissue handling, appreciating the operative steps, offers the facility for constructive feedback and poses no threat to patient safety.

A proficiency based progression-training (PBPT) curriculum requires achievement of expert-derived performance goals. This method of training enhances motivation, which is aimed to maximise skill acquisition and retention (Palter et al., 2013). It tailors training to meet the trainees individual needs while simultaneously ensuring that all trainees reach the same endpoint in terms of surgical performance. (Brunner et al, 2004; Korndorffer et al, 2005; Stefandis et al, 2005). It essentially allows trainees to acquire skills at different rates.

In order to provide the ideal proficiency based training model for surgical training there are a number of factors to consider. Firstly a structured curriculum needs to be developed (Aggarwal et al., 2007a, Walker M, 1998). Wiggens and McTigue's backward design approach to curriculum development for technical skills is one approach that has been proposed for surgical simulation (Wiggins G, 2001). Training should be carried out in a stepwise manner where the trainee begins on a simulator in the skills lab until predefined proficiency criteria are reached (Grantcharov and Reznick, 2008). An example of this is part-task training. Part-task training is a learning strategy whereby a complex task is deconstructed into smaller components for practice. Trainees gain proficiency in the individual components before progressing to the more complex task (Kolozsvari et al., 2011b). It is thought that a higher level of skill can be attained if participants master individual components before integrating them into the whole task.

There needs to be clear criteria to determine the competence level of the trainee and skill mastery (Sweet et al., 2010). The setting of training goals ensures that the trainee is required to reach a predefined standard and competence is not determined by time spent on the simulator or by performing a set number of repetitions. Standards should be benchmarked against both clinically established and simulator generated data. When this has been demonstrated and assessed in an objective manner then the trainee can progress to the real life operating room.

Training sessions should be spread out over a period of time in order to better augment and optimise learning (Moulton et al., 2006). Previously it has been shown that one hour on a virtual reality simulator equates to two hours spent in the operating room (Aggarwal et al., 2007b). In order for any training programme to be effective the virtual reality simulator needs to demonstrate acceptability, validity, reliability and reproducibility in the real life operating environment (Seymour et al., 2002). Evidence has demonstrated that PBPT improves intraoperative performance for surgical trainees (Grantcharov et al., 2004, Seymour et al., 2002).

1.6.2 Mapping Learning Curves

A learning curve is a graphical representation of the changing rate of learning (Figure 1.1). Typically the increase in retention of information is sharpest after the initial attempts. This increase gradually flattens out as less and less new information is retained after each repetition.

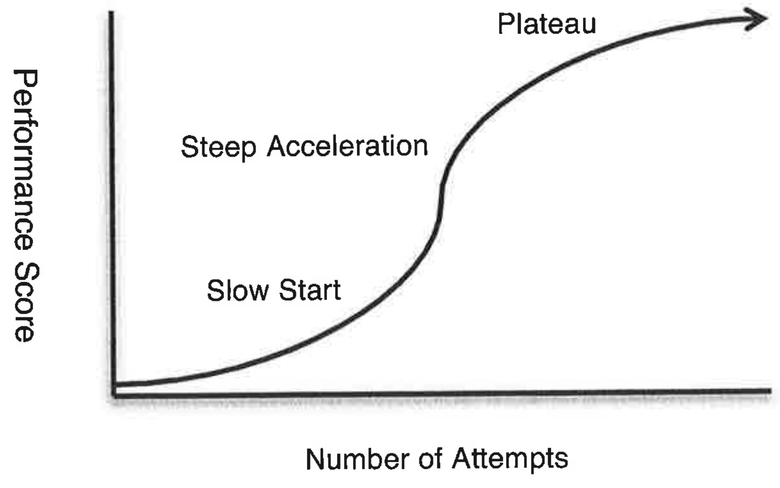


Figure 1.1 Graphical representation of the learning curve

As mentioned earlier, simulation provides a protected environment for trainees to overcome the initial learning curve. This concept has been discussed and examined by researchers in several studies over the last ten years.

Gallagher and Satava carried out a study (Gallagher and Satava, 2002) which looked at using the MIST-VR trainer as a tool for assessing psychomotor performance. As an adjunct to this, they also looked at learning curves. Both senior (<50 laparoscopic operations) surgeons and junior surgeons (<10 laparoscopic operations) performed six tasks on the MIST-VR; by trial 10 there was a convergence of mean performance. This showed that juniors could potentially perform to the level of a senior surgeon with practice outside the operating theatre.

A study by Grantcharov (Grantcharov et al., 2003) showed that different learning curves exist for surgeons with varying levels of laparoscopic experience. In this study, it was established that the MIST-VR was capable of differentiating between surgeons with different laparoscopic experience, which is, important for both construct validity and also for the potential development of internationally accepted norms of performance. If this was further developed then a trainee could use this as a reference point to establish where they currently are on the learning curve. Similar results were shown by Eversbusch (Eversbusch and Grantcharov, 2004). Three different learning curves were mapped for colonoscopy. The learning rate on the simulator was proportional to prior experience with endoscopy, which indicated that the simulator could assess parameters that are clinically relevant. Psychomotor training using the GI mentor compared with a control group who received no training demonstrated improved performance in the novice participants.

Aggarwal et al have produced several studies involving mapping learning curves. In a study in 2006 (Aggarwal et al., 2006a), two different learning curves were mapped out using medical students who performed tasks of various complexities on the MIST-VR. All three parameters (time, economy of movement and error scores) plateaued at the second repetition for the twelve core skills and at the fifth repetition for the most complex two tasks. Another study in 2006 (Aggarwal et al., 2006b) assessed the learning rate for dissection of Calot's triangle. A learning curve for novices was established as their performance plateaued at the fourth repetition. Learning curve data was established in a study in 2009 (Aggarwal et al., 2009) to ensure that repetitive practice improved performance, as measured by the simulator. Moreover by applying a stepwise process to learning a laparoscopic cholecystectomy, a whole procedure-based training curriculum could be developed. The learning curve for this procedure plateaued for all metrics between six and nine repetitions.

When laparoscopic suturing was examined in a study by Botden (Botden et al., 2009b), the number of repetitions required to reach the top of the performance curve (defined as proficiency) was eight knots. Lin et al (Lin et al., 2010) evaluated the learning curve for laparoscopic appendicectomy and found that operative duration and complication rate decreased in proportion to the increasing experience of the resident.

Interestingly, Grantcharov's study in 2009 (Grantcharov and Funch-Jensen, 2009) assessed the learning curve patterns of acquisition of generic skills in laparoscopy. In this study it was hypothesised that the familiarization rate with laparoscopic technique is different depending on psychomotor ability. Four types of learning curves were identified, proficiency from the beginning (5.4%), ability to advanced with practice which was found to be between two and nine repetitions (70.3%), ability to improve but unable to reach proficiency (16.2%), and finally no tendency to improve and overall underperformance (8.1%). This data suggests a role for developing a proficiency-based curriculum based on innate psychomotor ability. Several studies have looked into aptitude tests, which may relate basic laparoscopic technical skill performance (Gallagher et al., 2003, Hassan et al., 2007). Further to this research has attempted to ascertain the rate of skill acquisition in relation to innate ability (Stefanidis et al., 2006).

1.7 Role of Aptitude in Laparoscopic Surgery

1.7.1 Role of Aptitude in Learning Laparoscopic Techniques

Most of the difficulties encountered when learning laparoscopic surgery can be explained by visual spatial, psychomotor and perceptual factors (Gallagher et al, 1998; Perkins et al, 2002). One of the main visual spatial and perceptual problems is that the surgeon is required to interpret a 3-D image from a 2-D monitor. This is further complicated by a reduction in binocular information, as no clues are available to the depth of the objects being displayed. Normal binocular vision gives important information on depth perception, which is lost in laparoscopy due to the single point perspective.

The various visual spatial difficulties encountered during MIS are related to cognitive mapping and hand-eye coordination problems. Visual spatial discrepancies are also caused by a misinterpretation of the angular relationship, as the entry points of instruments do not correspond with the optical axis of the camera. These difficulties make accurate planning and executing movements within the abdomen more complex. (Crosthwaite et al, 1995). The reduction in the tactile sensation of the hands can cause difficulties in delicate surgical procedures and also results in the loss of ability to diagnose tissue using sensory judgment. The reduction in tactile feedback can result in difficulty navigating through the abdomen during laparoscopic surgery.

One of the main fundamental ergonomic problems associated with MIS is that instrumental manipulation is limited to only four degrees of freedom as opposed to six in open surgery. The fulcrum effect is considered one of the greatest problems limiting the surgeon's ability to acquire psychomotor skill in laparoscopy (Crowthers et al, 1999). The fulcrum effect is defined as the perceived inversion of movements. Laparoscopic surgery creates a visual discordance between the eye and proprioceptive information. The resultant feedback becomes counterintuitive causing an incorrect sequencing of psychomotor output, which requires a long period of time to overcome. (Gallagher et al, 1998).

1.7.2 Aptitude Assessment for Minimally Invasive Techniques

Aptitude assessment relies on the ability to demonstrate differences in desirable innate qualities that are essential to a specific task (Annett, 1974). Skills displayed by an individual while operating represent an individual's ability to adapt to MIS but provide little insight into the innate factors that mediate the level of performance.

Psychometric or mental measures are the sound and accurate measure of differences in individuals (Cooper, 1998). The aim of administering these tests is to gain an understanding of individual's attributes and to try and predict future performance (Edenborough, 1994). Psychometric tests originated in work related to education in the 1800's (Cattell, 1890). They were initially employed in the First World War (Yoakum & Yerkes, 1920). Following the war tests were developed to assess suitability for certain employments. Some examples include aptitude tests for clerical work (Burt, 1922) and for dress-making (Spielman, 1923).

The first battery of aptitude tests was published in 1947 in the US and was called the Differential Aptitude Test (DAT). This was a multi-occupational guidance package comprising of a range of tests covering abstract reasoning, spatial relationships and spelling.

The accurate identification of those individuals' best suited to an advanced career course or which job applicants would best suit a position can bring significant financial and personal benefits to an individual and company whilst reducing potential problems (Cooper, 1998). Aptitude is synonymous with ability and both refer to stable and innate characteristics such as intelligence and manual dexterity that represent the individuals level of cognitive processing and skill when dealing with a particular task.

The aptitudes, which have shown to be important for minimally invasive surgery, are visual spatial, psychomotor and depth perception. Visual spatial disorientation and misrepresentation have been implicated as potential causes of error during laparoscopic surgery (Hassan et al., 2007, Hedman et al., 2006). Further to this visual-spatial ability has shown to correlate with successful surgical performance in medical students (Wanzel et al., 2003). Psychomotor aptitude and its relevance to good performance in all surgical aptitudes and in particular minimally invasive surgery has been shown in the literature (Stefanidis et al., 2006).

All of the tests, which have been used to measure aptitude in the experiments in this thesis, have good test-retest reliability measures, which imply stability over time (Dikmen et al., 1999, Ekstrom RB, 1976).

1.7.3 Correlation between Aptitude and Laparoscopic Performance

Previous work in the National Surgical Training Centre in RCSI has established a significant association between visual spatial, depth perception and psychomotor aptitude for laparoscopic surgery (Nugent, 2011). Further to this, when the impact of psychometric aptitude on the ability to reach predefined proficiency goals was examined, it was found that those with a higher aptitude reached proficiency faster therefore a high aptitude shortens the steepest part of the learning curve. An important point noted during this study was that all subjects improved during proficiency based training of basic skills, even those with lower aptitude scores; however they needed more time committed to their training.

1.8 VR-to-OR Transfer

It is intuitive that training in a simulated surgical setting implies improved skill in a clinical environment. However this important concept requires definite clarification. There is little value to developing sophisticated training programmes in a simulated laboratory if laboratory training does not improve clinical performance. Transferability is often called VR-to-OR (a term coined by Professor Anthony Gallagher) and refers to the ability of simulation-based training to improve clinical performance. Transferability in clinical terms would imply predictive validity which is the capacity to improve future performance.

Such trials are usually designed by using two groups who are randomised to either receive simulation based training or no training. Their performance is then compared in a specific laparoscopic procedure or task after simulation training or no training. The groups ideally have similar baseline psychomotor and visuospatial ability. Assessment in the operating room is performed by an examiner who is blinded to the status of the subject, using the methods described previously. Even with sound methodology human trials can have many logistical challenges therefore many investigators opt to conduct their trials using animal specimen's most commonly porcine models. Clinical transferability can be shown with animal models in suitable laboratories as a bridge to the human setting. Transferability studies are essential in order to assess the ability of simulation based training to improve surgical performance in the operating room. They also require approval from an institutional review board.

If a study evaluates performance in a clinical setting after training in a skills laboratory and correlates the two performances, then predictive validity is supported which is the ability of a test the predict future performance in a different setting. If a study uses a validated rating tool as a gold standard to correlate the performance in an operation room to that in a skills laboratory, then concurrent validity is demonstrated.

In simple terms, the overall aim of transferability trials is to ideally detect differences in operative performances following simulation skills training. The key to being able to demonstrate the effectiveness of simulation based training in improving performance in the real life clinical setting is to use precise methods of assessment. The introduction of VR simulators has brought about huge advancements in surgical education as a surgeon's progression can now be measured in an objective way. Simulators are essentially computers, which can generate a profile summary upon completion of a procedure or a task, which provides immediate feedback and performance metrics.

The problem with correlating this to clinical performance is that the human body cannot act as a computer and provide such feedback. Therefore sound subjective assessment must be used in marking the live performance and after the intervention of simulation training in order to demonstrate the effectiveness of such training.

The first study to demonstrate a transfer of simulator learned skills to the operating room was in Yale, 2001 (Seymour et al., 2002). The control group had no simulation training and the trained group was taught to proficiency under supervision with emphasis on avoidance of errors. Candidates were assessed on dissection of the gallbladder from the liver edge both pre and post training or no training in the operating room during human cholecystectomies. The scoring system used was a novel pre-defined eight-error checklist; occurrence of these errors was recorded during each minute of the assessment. This was used instead of a global rating scale in an attempt to determine errors more accurately. A non-significant difference was detected in dissection time, with the trained group removing the gallbladder 29% faster than the non-trained group. In relation to error performance, the control group were five times more likely to burn the liver edge or injure the gallbladder and nine times more likely to fail to progress. Further evidence which supported this landmark research was in a study by Grantcharov et al (Grantcharov et al., 2004) which assessed both a trained and a control group in the clipping and cutting of the cystic duct. Again both groups underwent pre and post testing in the operating room during human cholecystectomies. Performance was measured using a modified OSATS scale by combining traditional parameters to create new parameters. It was found that the group who received simulated training on the MIST-VR performed faster, had greater economy of movement scores and lower error scores than the control group in the post-test assessment in the operating theatre, hence the study demonstrated transferability.

Following on from this initial research, various other studies demonstrating transfer of skill have been published. Some of them have shown partial task transfer and some using whole laparoscopic procedures, the latter of which laparoscopic cholecystectomy form's the bulk. Three other studies (Ahlberg et al., 2007, McClusky et al., 2004, Scott et al., 2000) assessed the transfer of skill in laparoscopic cholecystectomy. Scott used OSATS and demonstrated a significant improvement in the trained versus control groups. McClusky and Ahlberg used total error scores; both studies showed that error scores were higher in the control groups.

Other studies have looked at the transfer of whole procedures. A study by Larsen et al (Larsen et al., 2009) assessed the performance of an entire laparoscopic salpingectomy using an OSATS scoring system and found significant differences between trained and control groups. These same results have been shown with both laparoscopic hernia repairs (Hamilton et al., 2001) and laparoscopic nephrectomy (Traxer et al., 2001). When laparoscopic appendicectomy was assessed on a porcine specimen, the results of this study did not show any difference between trained and control groups. In this study, training time was very short, with three hours training in total. Achieving proficiency in a shorter time frame may have been difficult and therefore could have affected the outcome of this study. The assessment method used was blinded rater analysis using a scale of bad, average and good, which had no previous validation in this setting.

One study (Lucas et al., 2008) provided training for the novice group in laparoscopic cholecystectomy but assessed skill transfer in laparoscopic nephrectomy. The results showed that the group who received time based simulated laparoscopic cholecystectomy training outperformed the control group when a laparoscopic nephrectomy was performed in a porcine model. The students were assessed using OSATS. This shows not only the transfer of skill after simulated training but also that specific skills learnt for certain laparoscopic procedures are useful for other laparoscopic procedures.

Laparoscopic tasks as well as laparoscopic suturing have also been explored. Three studies (Korndorffer et al., 2005, Stefanidis et al., 2008, Verdaasdonk et al., 2008) evaluated the transfer of laparoscopic suturing. Two of them (Korndorffer et al., 2005, Stefanidis et al., 2008) used a formula which was 600 – [(time + (10 x accuracy score) + (10 x security error)]. This method awarded higher scores for the most accurate performance in the faster time. The purpose of this formula was to establish one value which if high implied a fast accurate performance and a good quality knot. By assigning one value to the user's performance as opposed to three, it gives results that are easy to compare and understand. Both studies showed significant improvements in the trained group compared with the control group. The third (Verdaasdonk et al., 2008) study used an error scoring system which showed that the control group made more errors than the trained group and this study also performed blinded rater video analysis looking at economy of movement and error assessment. Verdaadonk et al did not show any significant difference in the transfer of skill between the simulation-trained group and the control group.

It is apparent that simulation provides a safer more controlled environment of learning surgical skills and entire surgical procedures. The daunting task, which several randomized control trials have undertaken in order to prove the value of simulation-based training in the clinical setting, has furthered this concept.

1.9 Objectives

1.9.1 Hypothesis underlying the Objectives

The principle area that we planned to investigate was the role of aptitude on surgical performance. Our objective was to evaluate the impact of innate ability upon the rate at which a novice trainee can achieve proficiency in index and advanced laparoscopic procedures. We hoped to provide evidence that would guide future selection of trainees whereby those with higher innate ability could be candidates for fast-track training while those with lower innate ability may require a more intensive training pathway. The secondary focus of this thesis was to develop novel objective assessment methods of laparoscopic performance.

We know that in the current training climate, a structured curriculum using proficiency-based progression training methods is optimal. We therefore wanted to establish the learning curves that trainees must overcome in order to become proficient in MIS and determine the impact innate ability has on the length of these learning curves. Further to this, we wanted to develop new appropriate objective methods of technical skills assessment in order to aid the NSTC as a training body to assess trainees appropriately during the surgical training pathway. We hypothesised that by providing meaningful scores for the trainees, they would understand their performance scores in relation to their target goals and hence receive relevant feedback in relation to their performance. We aimed to validate these new scoring methods, which would aid an optimal proficiency-based training programme.

1.9.2 Detailed Objectives

Objective 1. To investigate the impact that innate ability has on completion of the learning curve for a basic laparoscopic procedure as well as a complex laparoscopic task.

A significant association has been found between performance in laparoscopic surgery and visual spatial, depth perception and psychomotor aptitude. It has also been found that those with a higher aptitude reached proficiency faster therefore a high aptitude shortens the steepest part of the learning curve. Our objective was to establish the impact of aptitude on the rate of achieving proficiency in laparoscopic appendicectomy and laparoscopic suturing in two groups of medical students with grossly different aptitude scores by comparing the number of attempts required to achieve proficiency.

Objective 2. To evaluate the role of self-selection into surgery based on fundamental ability.

Previous work in our institution raised the possibility of self-selection into surgical training based on fundamental ability. Our objective was to explore this further by comparing the visual spatial, psychomotor and perceptual aptitude scores of general and plastic surgery higher surgical training (HST) applicants to a group of medical students (surgical novice group) who were interested in pursuing a career in surgery. We also wanted to investigate if there was a proportion of (higher surgical trainees) HST's who had lower aptitude scores than the mean scores of the surgical novice group.

Objective 3. To provide construct and concurrent validity for a novel zone metric used to assess laparoscopic suturing.

In the setting of the aforementioned challenges that our training body faces, structured training is essential. Trainees need performance goals set by experts which are meaningful validated methods of assessing their progress. We planned on developing a novel objective method of assessing a complex laparoscopic skill such as laparoscopic suturing.

Objective 4. To establish a method of comparing laparoscopic performance scores of surgical trainees to expert surgeons by developing a novel objective scoring system for the ProMIS simulator

One of our aims was to develop a objective scoring system which would provide meaningful metrics and feedback of performance in relation to experts for surgical trainees.

In summery, the primary aim of this thesis was to evaluate the impact of aptitude (visual spatial, perceptual and psychomotor) upon the rate at which a candidate becomes proficient in laparoscopic procedures (laparoscopic appendicectomy and suturing). Our secondary aim was to validate novel objective assessment methods of laparoscopic performance by providing construct and concurrent validity for a novel zone metric used to assess laparoscopic suturing. We also hoped to develop a method of providing meaningful metrics scores in comparison to experts for the ProMIS simulator. Finally we planned to evaluate the role of self-selection into surgery based on fundamental ability.

Chapter 2

Materials and Methods

2.1 Ethics

Ethical Approval was granted by the Research Ethics Committee, Royal College of Surgeons Ireland under REC490 and REC558. All volunteers who participated in the experiments were given an information sheet to read describing the purpose of the research study and what would be expected of them during the course of the study. Once the subjects understood all the information provided and asked any questions they had, they were asked to sign a written consent form in accordance with the protocol detailed in the ethics proposal. All data collected was stored anonymously in excel spreadsheets with coded subject identification numbers. It was made clear to all the subjects that the data collected was used for research purposes only and would not be redistributed or shared with any third parties.

2.2 Participant Recruitment

An email was circulated to all the medical students in all years (1-6) within the RCSI by student services. All those who volunteered to participate were asked to undergo psychometric testing prior to being chosen for participation in the research study. This was done in order to establish two groups of medical students with baseline differences in innate ability. Participation was voluntary in all experiments and selection of participants was based on their aptitude scores.

For the basic surgical trainees (BST's), an email was sent to years 1 and 2 by myself the prinicipal investigator with the permission of the surgical training office in the NSTC. Participation was voluntary and all candidates who volunteered were selected for inclusion in the research study. The BST's who attended the NSTC for their yearly skills assessments were also asked to consent to allow their performance scores to be used for research purposes. When the applicants to the higher surgical training scheme attended the NSTC for aptitude and skills assessment upon successful shortlisting, they were asked to consent to allow their skills performance scores and aptitude scores to be used for research purposes.

2.3 Participant Demographics

In total, 132 medical students, 165 BST's, 154 HST's and 12 laparoscopic experts participated. Specific demographic details pertaining to each experiment are included in the methods section of each chapter.

All medical students in the pre clinical years (1-3) were included in the study and students in the clinical years (4-6) were only included if they had gained no hands-on experience during any of their surgical attachments or electives. Therefore if they had assisted in any surgical procedures or carried out any surgical tasks, they were excluded from the study. This ensured that all medical students were medical students and that there would be no confounding factors when testing the hypothesis.

Inclusion criteria for the basic surgical trainees depended on the study involved. In order to map specific learning curves for basic surgical trainees, ideally the trainees needed to be inexperienced in performing laparoscopic procedures. They were excluded from the study if they had performed any laparoscopic procedures unsupervised. All higher surgical trainees were eligible for the study.

2.4 Aptitude Assessment

Aptitude is a measure of a person's ability to acquire a specific set of skills, through future training. It is one's inherent capacity, talent or ability for a given task or activity. Aptitude tests assume that people differ in their innate abilities and that these differences can be useful in predicting future learning ability. The three components of aptitude considered important for laparoscopic surgery are visual spatial aptitude, psychomotor aptitude and depth perception. The various aptitudes and their respective tests are summarised in Table 2.1.

Table 2.1. Definitions and tests of aptitude

Aptitude	Definition	Test
Visual Spatial Aptitude	The ability to generate, transform and retain structured visual images	
• **Spatial orientation**	*The ability to mentally rotate a configuration*	**Card rotation and cube comparison test**
• **Spatial visualisation**	*The ability to apprehend a spatial object and match it with another spatial object*	**Map planning test**
• **Spatial scanning**	*The ability to quickly and accurately assess a complicated spatial field or pattern*	**Surface development test**
Psychomotor Aptitude	The ability to perform motor tasks with precision and coordination	**Grooved Pegboard**
Depth Perception Aptitude	One's visual ability to perceive the world in 3-D	**PicSOr**

2.4.1 Visual-Spatial Aptitude

Visual-spatial aptitude is the ability to generate, transform and retain structured visual images. It is one's ability to mentally manipulate 2-D and 3-D figures. The specific domains that are useful in evaluating laparoscopic performance are spatial visualization, spatial scanning and spatial orientation. The three paper tests, which were used to test these attributes, have good test-retest reliability measures (Table 2.2). The tests chosen have previously shown to be good markers of the domains in question.

Table 2.2. Test-retest reliability of aptitude tests

Aptitude Test	Test-retest Reliability
Card rotation test	0.80 (Ekstrom RB, 1976)
Cube comparison test	0.84 (Ekstrom RB, 1976)
Map planning test	0.80 (Ekstrom RB, 1976)
Surface development test	0.90 (Ekstrom RB, 1976)
Grooved Pegboard	0.82 (Dikmen et al., 1999)
PicSOr	0.94 (Crothers, 2001)

2.4.2 Visual-Spatial Aptitude Assessment

The kit of factor-referenced cognitive tests (1976) contains seventy two marker tests which are used to identify twenty three aptitude factors (Stylopoulos et al., 2004). Four visual spatial paper-based tests were selected from this kit. These included the card rotations test, cube comparison test, map planning test and surface develoment test (Appendix III).

The tests are administered in a standardised fashion. The first page is an instruction page for which two minutes are given to read. Included in this page are a number of examples that the subject can practice within this two minutes. The test contains two parts, part A and part B. Three minutes is allocated for each part. The subject is asked to complete as many questions as possible for each part without sacrificing accuracy.

2.4.3 Spatial Orientation Assessment

Spatial orientation is one's ability to maintain orientation in relation to the surrounding environment. It is the ability to mentally rotate a configuration.

Spatial orientation is assessed using the card rotations test and the cube comparison test. Both of these tests are negatively marked which should be highlighted before the test begins. The score for both these tests is calculated by adding up the number of correct answers minus the number marked incorrectly.

Card rotations test (Figure 2.1) assesses ones ability to mentally rotate a 2D object in space. There is a total of 160 questions in this test.

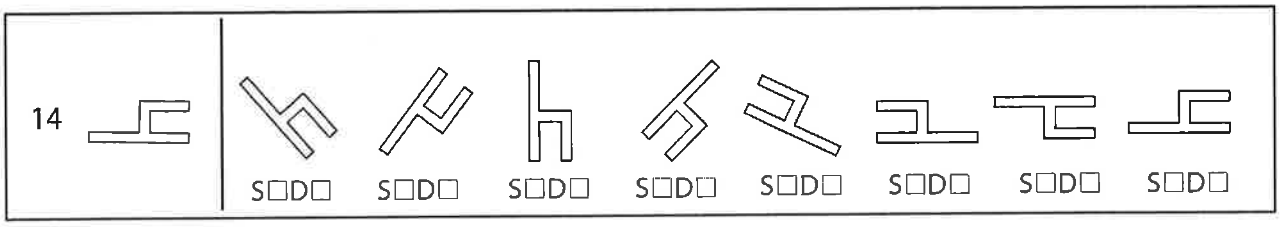


Figure 2.1 Card rotations test Each problem in this test consists of one card of the left of a vertical line and eight cards on the right. The subject must decide whether each of the eight cards on the right is the same as or different from the card on the left. Underneath each card are boxes labelled either S or D. The subject is instructed to mark the box S if it is the same as the one at the beginning of the row or to mark the box D if it different from the one at the beginning of the row.

The cube comparisons test (Figure 2.2) assesses ones ability to rotate a 3D object in space therefore it is a more difficult version of the cards rotation test.. There is a total of 42 questions in this test.

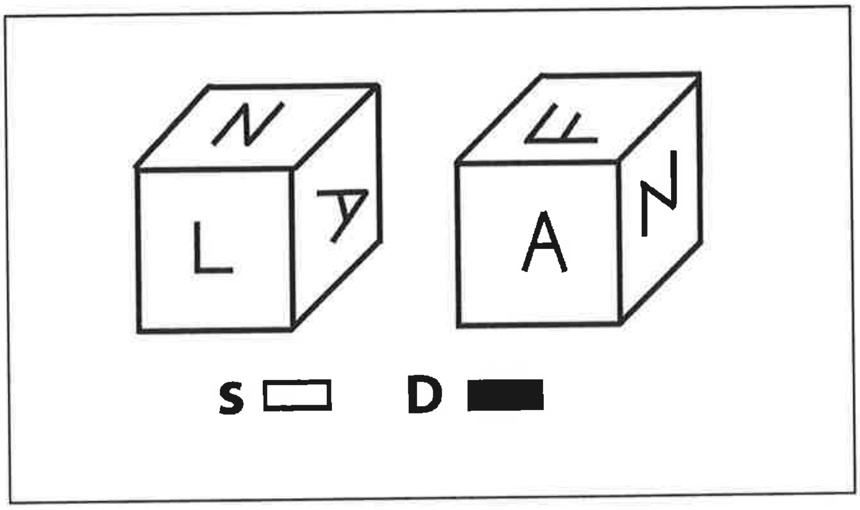


Figure 2.2 Cube comparison test. Each problem in this test consists of drawings of pairs of cubes The cubes represent wooden blocks which may have a different letter, number or symbol on each of the six faces (top, bottom and four sides). As with the cards rotations test, each of presented pairs of cubes have a box labelled S and a box labelled D. The subject must decide whether the pair of cubes are the same or different. The subject is instructed to mark the box S if they are the same or to mark the box D if they are different. A very important point with must be highlighted to the subjects is that no letter, number or symbol can appear on more than one face of a given cube. Except for that, any letter, number or symbol can be on the hidden faces of a cube.

2.4.4 Spatial Visualisation Assessment

Spatial visualization is one's ability to apprehend a spatial form or object and match it with another spatial form or object. Only parts of the figure are manipulated unlike with spatial orientation where the whole figure is manipulated. Spatial visualization is more complex than spatial orientation.

Spatial visualisation was assessed using the surface development test. This test asseses ones ability to create a 3D object from a 2D object in space. The subject is asked to imagine or visualise how a piece of paper can be folded to form some kind of object.

Each part contains 6 questions with five stems giving a total of 60 questions in the test. The score is calculated by subtracted a fraction of the incorrect letters form the total of correct letters.

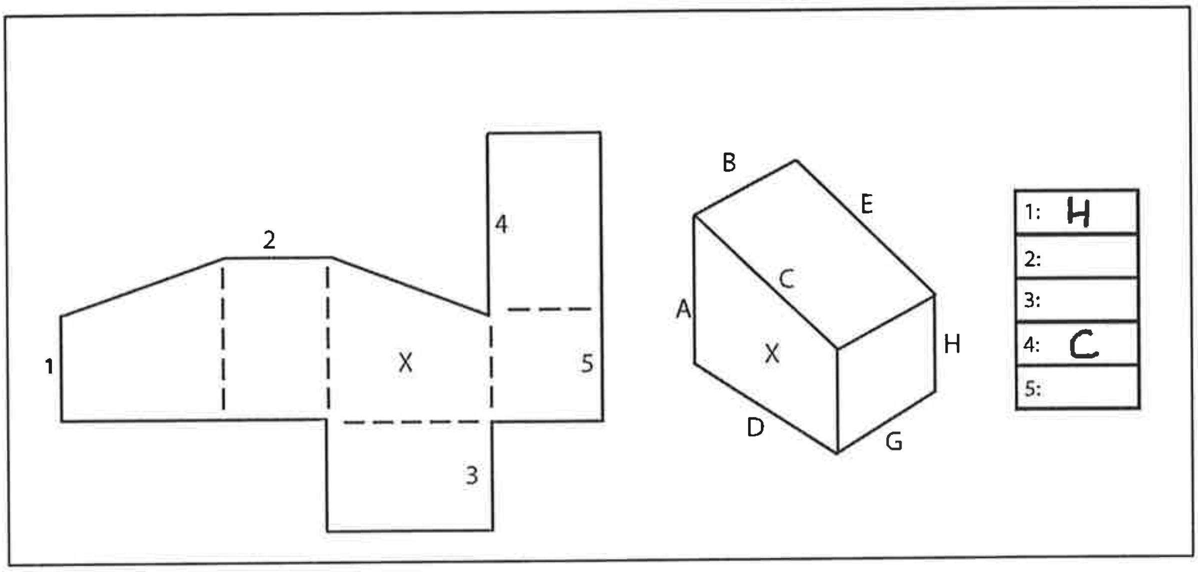


Figure 2.3 Surface development test. Two drawings are presented. The first drawing is a piece of paper which can be folded on the dotted lines to form the object depicted in the second drawing. The subject must imagine how the flat piece of paper would be folded to form the object. The flat piece has the numbers one to five marked on the sides with solid lines. The object presented has each side labelled with a letters from A to G. The subject must correspond with the number of the side on the flat piece of paper to the letter on the side of the object. Both the flat piece and the object have a face marked with an X. The side of the flat piece marked with an X will always be the same as the side of the object marked with an X. Therefore, the paper must always be folded so that the X will be on the outside of the object. It should be pointed out to the subjects that two of the answers can be the same per question.

2.4.5 Spatial Scanning Assessment

Spatial scanning is one's ability to quickly and accurately survey and assess a complicated spatial field or pattern in order to identify the correct path through the visual field. Spatial scanning was assessed using the map planning test. This test assesses ones ability to rapidly scan a complex visual field and find a pathway through it. There is a total of forty questions in this test. This test is not negatively marked.

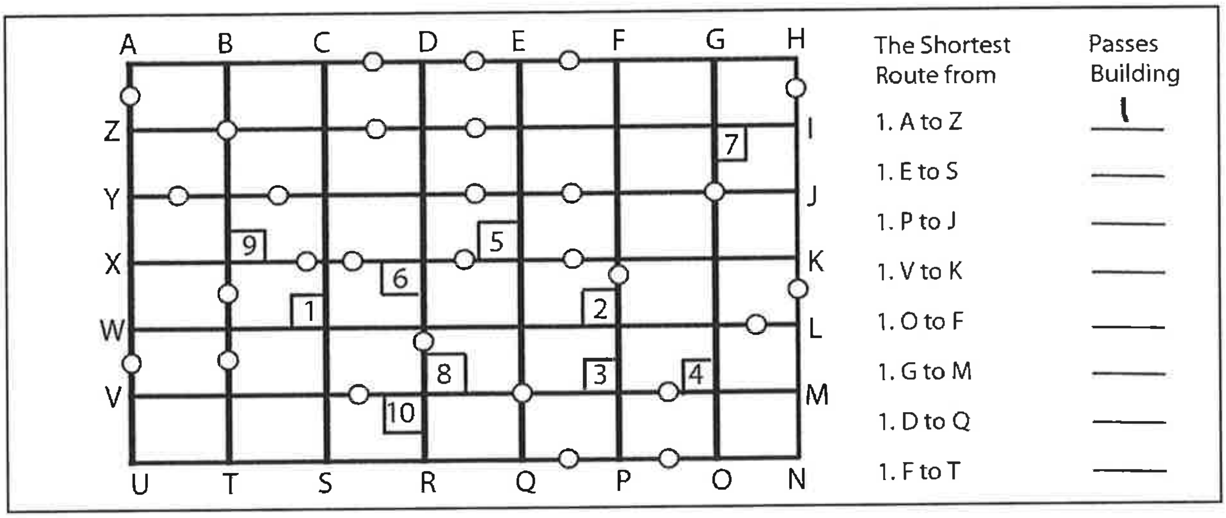


Figure 2.4 Map planning test. A drawing of a map of the city is provided. The dark lines are streets and the numbered squares are buildings. The circles are road blocks which cannot be passed through. The subject is asked to find the shortest route between two lettered points. The number on the building passed is your answer. Certain rules must be highlighted prior to commencing the test. The shortest route will always pass along the side of one and only one of the numbered buildings, if the subject passes two buildings they have not gone the shortest route. A building is not considered as having been passed if a route passes only a corner and not a side. The same numbered buildings may be used more than once.

2.4.6 Psychomotor Aptitude

Psychomotor aptitude is the ability to perform motor tasks with precision and coordination. Specifically for this study we are interested in measuring manual & finger dexterity and hand-eye co-ordination. Manual dexterity is the ability to perform coordinated movements using one's hand and arm. Finger dexterity is the ability to perform co-ordinated movements using one's fingers. Hand-eye coordination is the ability to coordinate control of eye movement with hand movement in order to execute a task. It also involves the processing of visual input to guide reaching and grasping along with the use of proprioception of the hands to guide the eyes. Psychomotor aptitude was assessed using the Grooved Pegboard. It has been well validated and is commonly used. Test-retest reliability for the grooved pegboard has been reported as $r > 0.82$ (Dikmen et al., 1999).

2.4.7 The Grooved Pegboard

The grooved pegboard is a manipulative dexterity test. The unit (Lafayette instruments, model 32025®) is a metal board with a metal surface measuring 10 x 10 cm. It consists of 25 holes with randomly positioned keyhole shaped slots. A round area is located above the metal board which holds the pegs. Pegs must be rotated to match the hole before it can be inserted into the slot which has a key along one side. Visual input is gained from viewing the different keyhole shaped slots in the board. Fine motor dexterity is used when inserting the peg to match the orientation of the slot. This test requires complex visual-motor coordination which is not found in other motor tasks such as the Purdue Pegboard Test.

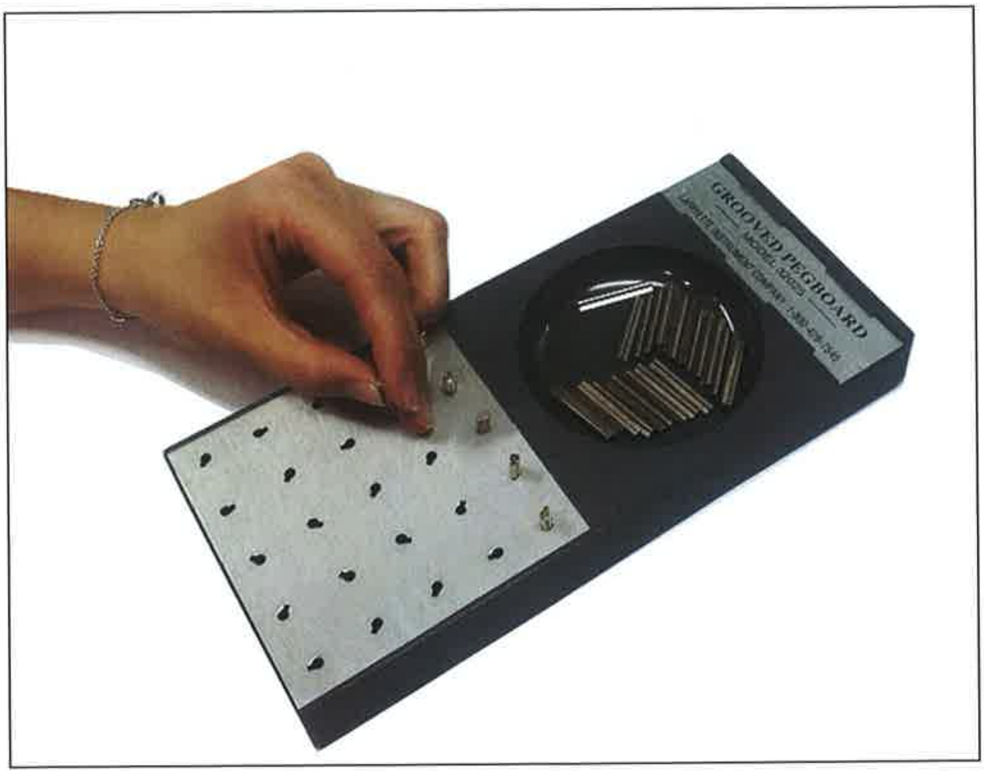


Figure 2.5 The Groved Pegboard (Lafayette instruments, model 32025). The instructor points out the pegs and the pegboard and explains that all the pegs are the same and that each peg has a square side and a round side that matches with the slots in the pegboard. The subject is instructed that the aim is to put all the pegs into the pegboard as quickly as possible using only one hand and without dropping any pegs. The pegs must be put in the board in the correct order and in the correct direction. Only one peg is to be picked up at a time and the subject should immediately be told if more than one peg is picked up. If a peg is dropped, the subject should continue using a different peg and not attempt to pick up the dropped peg.

2.4.8 Assessment using the Grooved Pegboard

The pegboard is placed in the center of the subjects visual field. Only one hand is used per attempt. Each hand is tested starting with the subjects dominant hand. For the right hand trial, the pegs are placed from left to right and for the left hand trial, the pegs are placed from right to left. The examiner should encourage the subject to perform the task as quickly as possible.

2.4.9 The Grooved Pegboard Scoring

The length a time in seconds required to perform the task is recorded. The task begins when the subject starts and ends when the last peg is inserted. The number of dropped pegs per trial should also be recorded. For each hand, the time score added to the number of drops made in order to get the complete score per attempt. An average score of the dominant and non dominant hands was then calculated in order to determine a score which reflects psychomotor performance.

2.4.10 Depth Perception Aptitude

Depth perception is one's visual ability to perceive the world in 3-D. A test called PicSOr (Pictorial Surface Orientation) was developed in order to assess a subject's perceptual ability in laparoscopic surgery (Cowie, 1998).

2.4.11 Pictorial Surface Orientation (PicSOr)

PicSor is where the surface slant of an object is used to describe the orientation of the object. The surface slant describes the angle between one of the axes of a surface of an object at the arbitrarily selected point, and its projection in a picture. PicSor tests the ability of the subject to recognise the equality of the slant for each object. It has been validated as an objective psychometric test in minimally invasive surgery. The test-retest reliability of PicSor has demonstrated to be r = 0.94.

2.4.12 Assessment using PicSOr

This test is performed using a software package called PicSOr (Figure 2.6). The candidate is given clear instructions as to the aim of the test and how to complete the required task. The candidate begins in practice mode. Eight attempts are practiced using a feedback box until the subject is happy to proceed. This allows the candidates to better understand the task being asked of them. Once the initial trials attempts are completed, the subject moves on to the experiment mode where there are required to complete 35 items. The experiment mode follows the same format as the practice mode but without the feedback box. There is no time limit to this test.

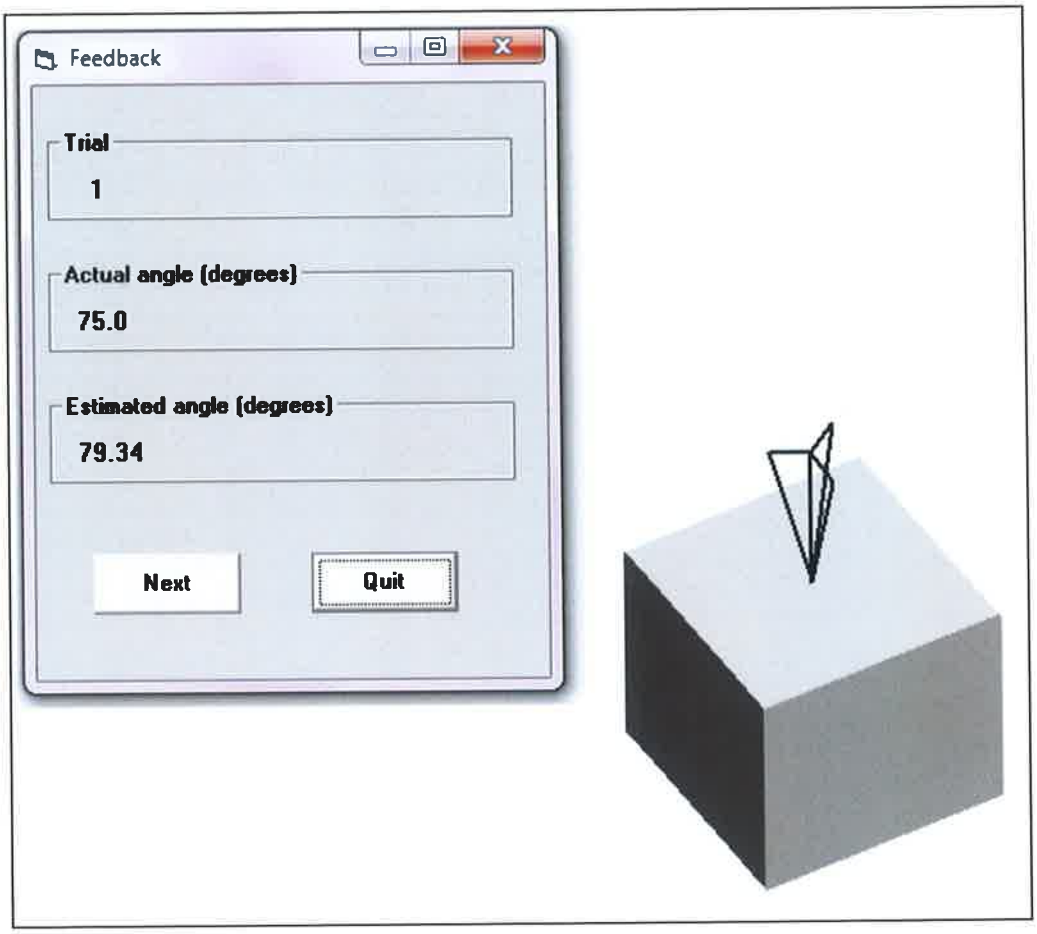


Figure 2.6 Pictorial Surface Orientation Test (PicSOr). When the programme begins, a screen appears with a 3D geometric cube and a spinning arrow top touching the surface of the cube. The task involves moving the arrowhead in order to create a 90 degree angle between the arrowhead tip and the surface of the cube. Once the subject is happy with the angle created and presses enter, a feedback box will appear. Two figures appear, the actual angle – the correct answer and the estimated angle which was guessed by the subject.

2.4.13 Scoring PicSOr

The closer the estimated angle figure is to the actual angle figure, the more accurate the answer. The final resting position of the arrowhead which is the estimated angle figure was correlated to the theoretically correct arrowhead orientation using microsoft excel. By correlating these two figures for the array of 35 items, a correlation figure was calculated which was used as a score for depth perception ability.

2.5 Surgical Skills Assessment

Laparoscopic skills were assessed using the ProMIS simulator. All assessments took place in the NSTC in the RCSI, 121 Stephens Green, Dublin 2.

2.5.1 ProMIS

The ProMIS 3 Simulator (Haptica®, Dublin) (Figure 2.7) was used for all assessments. It is a hybrid simulator which uses augmented reality that overlays graphics onto a task performed on a physical exercise. ProMIS enables the user to use physical models, which ensures appropriate tactile feedback. It tracks the instruments, thereby producing a report of metric results, which provides immediate feedback on performance. Numerous studies have provided construct validity for this hybrid simulator (Broe et al., 2006, Gilliam, 2009, Neary et al., 2008, Ritter et al., 2007).

The ProMIS simulator consists of a laparoscopic interface, which consists of a torso-shaped mannequin connected to a laptop where graphics are displayed. The mannequin contains three separate tracking cameras, arranged to identify any instrument inside the simulator from three different angles. The left and right tracking cameras are positioned to capture instrument motion looking in a caudal direction of the left and right sides of the mannequin, respectively. The central tracking camera is positioned at the mannequin's pubic symphysis looking cephalad and serves as the main viewing camera displayed on the computer screen. The camera tracking system captures instrument motion with Cartesian coordinates in the x, y, and z directions at an average rate of 30 frames per second. During simulation, yellow tape was applied on all instruments at a standardised distance from the tip of the instrument. The tape enables the simulator to track the motion of the laparoscopic instruments. Instrument movements are recorded until they are removed from the mannequin.

Figure 2.7 The ProMIS Laparoscopic Simulator

2.5.2 Metrics

Upon completion of a performance or task, ProMIS simulator generates an immediate profile summary of objective measurements. These include operative time, path length and economy of movement. Operative time is the length of the procedure measured in seconds. Path length is the distance travelled by the instrument or the sum of deviations from a fixed point, and is measured in millimeters. Economy of movement is a smoothness measurement and is detected by changes in instrument velocity. It has no units and is purely a numeric value. These metrics are recorded for each instrument (right and left hand) during simulation. This provides valuable information regarding the contribution of both hands to a given procedure which is very important from an ergonomic standpoint The software stores instrument data based on the coordinates of their respective entry point, which allows it to eliminate false positives and discriminate left and right instruments. The entry plane of the body form skin is divided into a left and a right half-plane along the simulator midline. The instruments are labeled left or right based purely on their entry-point position, which is constant in time as long as the instrument is not removed and reinserted at a different place.

The measurement of metrics is a good objective assessment of skills. They are calculated as cost functions, in which a lower value describes a better performance. When instrument path length is applied to laparoscopy, this suggests operative focus and greater overall performance and experience.

Smoothness is a measure of the sudden changes in direction and acceleration. Sharp turns create a high value while smooth movement creates a low value. The lower the value the better the performance. A good score implies good purpose of movement.

2.5.3 Modules

The laparoscopic appendicectomy, laparoscopic suturing modules were the primary modules used during the course of the experiments and will be explained in further detail in the subsequent chapters. The object positioning and sharp dissection modules were also used in order to develop a standarised objective scoring system.

All modules consist of both teach and test me modes. (screen shots) The "teach me" mode gives both verbal and onscreen graphic instructions on how to carry out each step. It also contains short video segments where relevant in order to display the step being performed in the live setting. The "test me" mode the onscreen graphical cues are removed by the simulator still gives verbal instructions on the steps required for each stage of the procedure. In order to test the subjects knowledge of the procedure aswell as their operative technical skills, a "blank" module was used during assessments so that the subject would not be prompted as to the next step in the operation.

The object positioning module requires the operator to move beads from one pot to another on a pre-designed tray in the laparoscopic simulator. The simulator instructs the operator where the beads should be moved at each step. At the end of the task, the operator is required to move all the beads into a bag placed on the tray. There are three levels in this module which progress in complexity. The sharp dissection module has two levels of complexity and requires the operator to cut a straight line and then cut out a triangle on a fixed glove. The object is to cut the upper layer of the glove without perforating the layer underneath. A record of the number of perforations is noted. The simulator will verbally instruct the operator at each step.

The laparoscopic appendicectomy module (Figure 2.7) was used in the demonstration of how to perform the procedure prior to assessment. The module contains 6 steps. This was used in teach me mode. A synthetic appendix model (Limbs & Things, Bristol, United Kingdom®) was inserted into the ProMIS simulator tray. In order to assess the candidates a "blank" module was preferred over the test me mode. Each candidate received adequate didactic teaching and watched both a live and simulated appendicectomy prior to assessment.

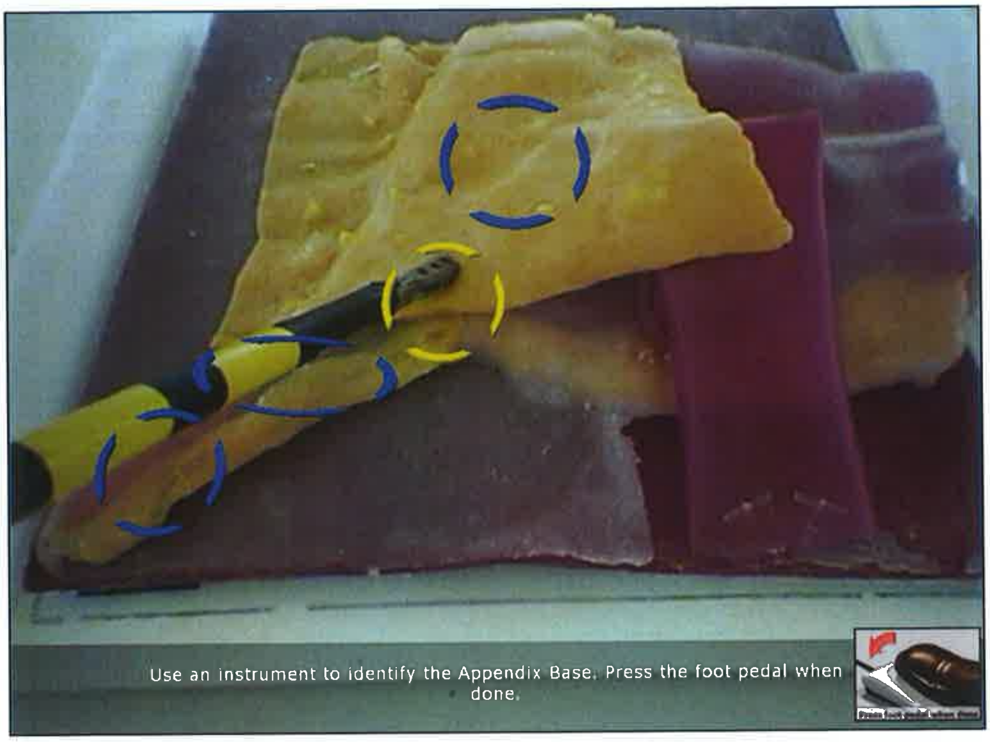


Figure 2.7 Laparoscopic appendicectomy module

The laparoscopic suturing module (Figure 2.8) was used in the demonstartion of how to perform intracorporeal suturing and knot tying. The module contains 5 steps. This was used in teach me mode. A 10 x 12 cm piece of synthetic suturing skin (The Chamberlain Group, Massachusetts®) was placed in the simulator tray. Again in order to assess the candidates a "blank" open module was preferred over the test me mode. Each candidate received adequate didactic teaching and watched a simulated performance of intracorporeal suturing prior to assessment.

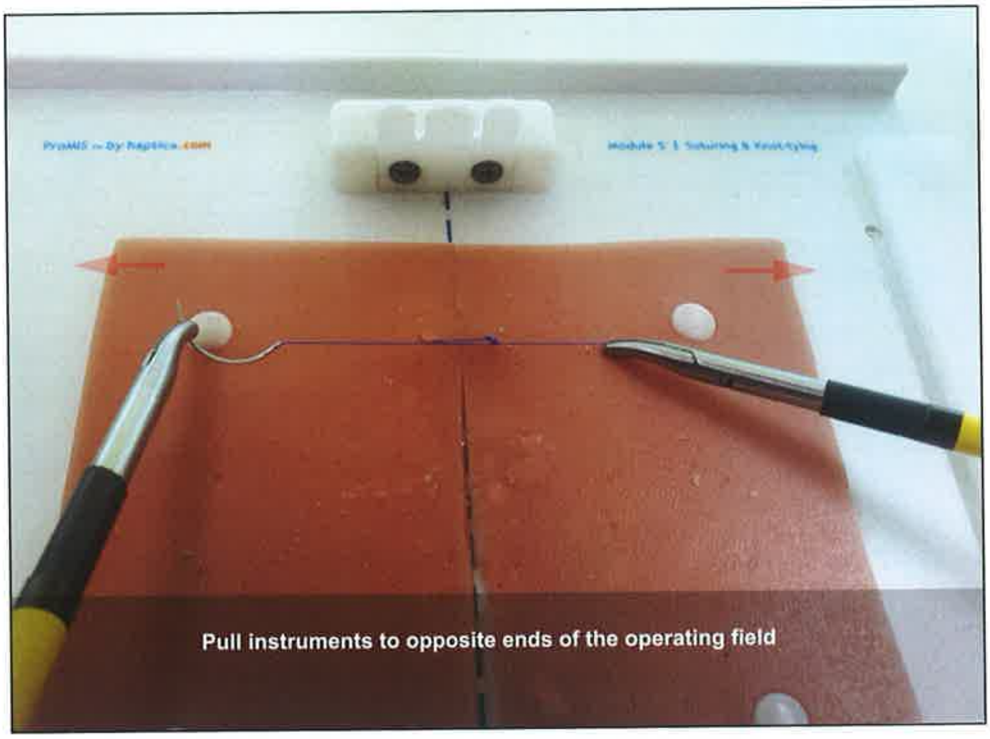


Figure 2.8 Laparoscopic suturing module

2.5.4 Subjective Scoring

Further to metrics, all the surgical performances were subjectively assessed using the Objective Structured Assessment of Technical Skill (OSATS) which was introduced by Reznick et al in 1993 and has high reliability and global validity. This seven item table of technical performance on a fivepoint grading scale includes respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, knowledge of specific procedure and use of assistants. Each performance was recorded and subsequently assessed by two reviewers, blinded to the experience level of the candidate. The laparoscopic suturing tasks were also assessed using the Fundamentals of

Laparoscopic Surgery (FLS) rating scale. The FLS program is accredited by the American College of Surgeons. This scale rates six key components of a laparoscopic suturing task on a five point grading scale (Appendix IV).

An effort was made to assess all performances subjectively as well as objectively in an attempt to appropriately validate metrics. We also aimed to provide the most valid assessment method possible.

2.5.5 Error Scoring

Each synthetic anatomy tray was examined after procedure completion for pre-defined errors, which will be detailed for each experiment in subsequent chapters. The error scoring was carried out by two assessors blinded to the experience level of the candiadate.

2.6 Statistical Analysis

The statistical package software Stata 12.0 was used for all analysis. Normality of data was assessed using the Shapiro-Wilk normality test. When it was acceptable to retain the null hypothesis, parametric tests were used. When it was deemed unacceptable to retain the null hypothesis non-parametric methods of analysis were used. The students t-test, wilcoxon rank-sum test and the k sample equality-of-medians test were all used for the comparison of two means depending on the normality of data and size of the dataset. The Wilcoxon signed-rank test was used to determine the significance of improvement in repeated performance scores. A p value <0.05 was considered statistically significant. Correlation coefficients were calculated to determine the strength of relationship between two continuous variables. Inter-rater reliability was determined using the intraclass correlation coefficient (ICC).

Chapter 3

Impact of Innate Ability on Completion of the Learning Curve for an Index Laparoscopic Procedure

3.1 Introduction

The various challenges currently facing surgical training programs have led to numerous changes in the way surgical training is being delivered. These challenges include reduced working hours, emphasis on minimally invasive techniques (which reduce the afforded opportunities for trainees to perform index procedures), as well as changing patient expectations with emphasis on consultant-delivered services. Although advantageous for the patient, the widespread application of minimally invasive techniques combined with the aforementioned challenges has created barriers to the traditional apprentice-model of surgical technical training. The skill set required in laparoscopy is very different compared with open conventional surgery (Figert et al., 2001). This is due to lack of tactile feedback, precise hand eye coordination, a change from 3-D to 2-D visualisation and adaption to the fulcrum effect (Gallagher et al., 1998, Perkins et al., 2002). Furthermore, the early part of the learning curve is associated with a higher complication rate (Moore and Bennett, 1995).

The widespread use of simulators has changed surgical training over the last decade (Aggarwal et al., 2007b, Gurusamy et al., 2009, Korndorffer et al., 2005, Seymour et al., 2002). There has been a shift towards implementing proficiency-based programs for surgical residents whereby one must demonstrate proficiency prior to progression (Ahlberg et al., 2007, Gallagher et al., 2005, Rosenthal et al., 2010). The attainment of technical competence is based upon completing the learning curve to a pre-established threshold set by surgical experts.

Aptitude is defined as a set of attributes that determine potential for a given activity. This potential may be developed into skilled behavior with training and practice. There are three main areas of aptitude that are considered relevant in minimally invasive procedures. These are visual spatial aptitude, psychomotor aptitude and depth perception.

Several studies have examined the relationship between specific areas of aptitude such as visual-spatial and perceptual ability with laparoscopic technical skill performance (Gallagher et al., 2003, Hassan et al., 2007). These studies have concluded that superior laparoscopic performance is demonstrated among novice surgical trainees who possess such attributes.

In 2009 Grantcharov et al (Grantcharov and Funch-Jensen, 2009) assessed the learning curve patterns in acquiring laparoscopic skills. This study concluded that the familiarization rate of laparoscopic techniques varies according to psychomotor ability and four types of learning curves were identified.

We have previously demonstrated that there is an association between aptitude and simulator performance of basic laparoscopic tasks and laparoscopic colectomy (Nugent et al., 2012b).

3.2 Objectives

3.2.1 Hypothesis

We hypothesised that a candidate's aptitude may have direct implications upon their ability to complete the learning curve for a minimally invasive index procedure. We wanted to quantify the impact of aptitude on attainment of proficiency in laparoscopic appendicectomy.

Our study was intended to identify subjects who may not improve with repeated attempts and fail to complete the learning curve for an index laparoscopic procedure such as laparoscopic appendectomy thereby providing objective measures to refine the selection process for future surgical trainees.

The primary aim of this study was to demonstrate the relationship between aptitude and the rate at which one becomes proficient in laparoscopic appendicectomy. By investigating the influence of aptitude of the rate of the learning curve, trainees who may require extra training in the simulation laboratory can be identified. The secondary goal was to design the expected learning pathway for surgical trainees for a basic laparoscopic procedure such as laparoscopic appendicectomy.

3.2.2 Detailed Objectives

Objective 1. To investigate the impact that innate ability has on completion of the learning curve for laparoscopic appendicectomy.

We aimed to compare the rate at which two groups of medical students with differing innate ability became proficient in laparoscopic appendicetomy.

Objective 2. To determine if medical students with high innate ability have superior baseline performance in laparoscopic appendicectomy than those with low innate ability.

In the same two groups of medical students, we aimed to assess their baseline performance by comparing metric, error and OSATS scores of the first laparoscopic appendicectomy performed between the two groups.

Objective 3. To map the number of attempts required by trainee surgeons' to reach technical proficiency in laparoscopic appendicectomy.

We aimed to evaluate the learning pathway of surgical trainees by establishing the number of attempts required to reach predefined goals. By doing this we aimed to determine the minimum number of procedures, which must be performed by trainees during their training.

Objective 4. To set benchmark proficiency levels for laparoscopic appendicectomy on the ProMIS simulator.

In order to map the learning curve for laparoscopic appendicectomy, we needed to establish proficiency levels. We aimed to recruit a group of expert consultant surgeons, who had performed greater than 150 laparoscopic appendicectomies.

3.3 Materials and Methods

3.3.1 Participant Recruitment

As outlined in chapter 2, Participants were recruited to take part in this study on a voluntary basis. Eighty medical students with no prior surgical experience were tested in the three different aptitudes described below. All participants were in years 1-3 of a 5-year medical school program.

Only first year surgical trainees were recruited as we hoped to recruit a group of surgical trainees with no or minimal laparoscopic experience as possible in order to map an appropriate learning curve for laparoscopic appendicectomy. Therefore any basic surgical trainee who had performed either a supervised or unsupervised laparoscopic procedure was excluded from the study.

Five laparoscopic experts was also recruited in order to set benchmark proficiency levels on the ProMIS simulator (Appendix V).

All participants who were selected were asked to sign a consent form allowing all data collected to be used for research purposes. It was made clear to all the subjects that the data was stored and presented in an anonymous format.

3.3.2 Participant Demographics

Based of the results of the aptitude tests, two groups of twelve were selected from opposite ends of the aptitude spectrum. The first group of twelve students (group A) were considered to have high aptitude with a score one standard deviation higher than the mean score of the study population. The second group of twelve students selected (group B) were considered to have low aptitude as their score was one standard deviation lower than the mean score of the study population. The students were blinded to which group they were in.

Twelve basic surgical trainees in the first year of their training were recruited to take part in the study. The inclusion and exclusion criteria for the study participants have been previously outlined in chapter 2.

3.3.3 Aptitude Assessment

The three main areas of aptitude considered to be relevant for minimally invasive procedures are visual spatial aptitude, psychomotor aptitude and depth perception ability.

Visual spatial aptitude was assessed using tests sourced from the Kit of Factor Referenced Cognitive Tests (Ekstrom et al, 1976). Four paper based tests were used; card rotations and cube comparison tests assessed spatial orientation, map planning test assesed spatial scanning and surface development test assessed spatial visualisation. Psychomotor aptitude was tested using The Grooved Pegboard which assessed manual dexterity and hand-eye coordination (Dikmen at all, 1999). Depth perception was assessed using a computer based software program known as Pictorial Surface Orientation (PicSOr) which tests ability to convert a 2-D image on a screen to a 3-D image (Cowie, 1998).

A more detailed explanation of these aptitude assessments and how they were carried out are outlined in chapter 2.

3.3.4 Setting Proficiency Levels

Five laparoscopic experts were recruited to set proficency levels. Each expert was asked to perform a laparoscopic appendicectomy on the ProMIS simulator using the same materials and under the same conditions as the medical students. They performed the procedure twice and a mean score of path length, smoothness and time was calculated for each expert. Proficiency was then determined by calculating the mean expert score.

3.3.5 Simulator & Materials

The ProMIS III® Simulator was used for assessment. Further details can be found in Chapter 2.

A synthetic appendix model (Limbs & Things, Bristol, United Kingdom®) was inserted into the ProMIS simulator tray. A new model was used for each attempt. The laparoscopic appendicectomy was performed in a conventional way using titanium clips to ligate the mesoappendicular vessels and endoloops to secure and transect the appendix base.

3.3.6 Surgical Performance Assessment

Prior to performance assessment each subject received didactic teaching. Each subject was sent a stepwise approach detailing how to perform the procedure (Table 3.1) and a video-link to a live recording of a laparoscopic appendicectomy before the experiment commenced. When the subjects attended the first session, a simulated laparoscopic appendicectomy was demonstrated. They were allocated time to ask questions before they attempted a mandatory multiple-choice questionnaire to ensure complete comprehension of the procedure. Prior to their first assessement, they had an opportunity to familiarise themselves with the testing equipment by completing a basic laparoscopic task. They were required to point to various points on the appendix model using the laparoscopic grasper.

Participants were asked to perform the procedure consecutively until they reached the proficiency scores. An interval practice curriculum as oppossed to massed practice was implemented (Gallagher et al., 2005, Metalis, 1985). Subjects were allowed to perform a maximum of three procedures per session to avoid fatigue (Verdaasdonk et al., 2007, van Dongen et al., 2011a) and sessions were spaced at maximum two weeks apart (Stefanidis et al., 2008). The subjects were supervised by a senior surgeon at all times and if the subject needed guidance, instructions were given however at no stage during any of the performance assessments did the senior surgeon take over the task. Upon completion of each procedure the simulator provided a summary report of the metric scores which were relayed to the subject so that they were aware of their progress throughout the experiment.

Each performance was recorded and subsequently assessed by two reviewers blinded to the status of the surgical novice using the OSATS scoring system (Faulkner et al., 1996). Each tray was examined after procedure completion for six pre-defined errors (Table 3.2) and this was also relayed to the candidates after each performance.

Table 3.1. Operative steps of laparoscopic appendectomy

1.	Check to ensure that the equipment is available and working
2.	Umbilical port is inserted using Hassan technique
3.	The 5 mm ports are introduced under vision (left iliac fossa and suprapubic areas)
4.	The caecum, terminal ileum and appendix are identified by touching them with a closed grasping forceps and naming them
5.	The tip of the appendix is grasped with the non-dominant hand and it is drawn anteriorly and inferiorly to create tension on the overlying peritoneum
6.	The peritoneum is divided with the scissors
7.	Put further tension on the appendix with the grasper in the non-dominant hand towards the right anterior superior iliac spine, this creates tension on the mesoappendix
8.	A space is created in the mesoappendix between the mesoappendicular artery and the appendix using the maryland in the dominant hand
9.	3 x clips are applied with the clip applier in the dominant hand to the vessels and the artery is divided between the more distal 2 clips
10.	An endoloop is introduced and the appendix is released from the grasper and then picked up through the endoloop with the nondominant hand
11.	The appendix is drawn towards the anterior abdominal wall with the grasper and the endoloop is opened and deployed close to the junction with the caecum
12.	The endoloop is then cut with the scissors in the dominant hand
13.	A second endoloop is then deployed (and cut) and the appendix is divided with a scissors distal to the two loops. It is removed through the trocar. *(In the interest of cost we will not use the third endoloop or a specimen retrieval bag)*

3.3.7 Statistical Analysis

A database was constructed using Stata 12.0. Data was analysed using descriptive statistics and non-parametrics tests. The aptitude, metric and proficiency scores of the two groups were compared using the Mann Whitney test. A p value <0.05 was considered statistically significant. Non-parametric testing was performed as none of the data was shown to have a normal distribution using the Shapiro Wilk test.

Table 3.2. Laparoscopic appendicectomy procedural errors

1.	Incorrect placement of ports
2.	Cutting the terminal ileum
3.	Cutting the caecum
4.	Incorrect placement of clips/ incorrect division of artery between clips 2 & 3 such that there are 2 clips on the specimen at the end
5.	Failure to place 2 endoloops on the appendix
6.	Failure to cut the appendix distal to the two endoloops

3.4 Results

3.4.1 Participant Demographics

In total, forty one subjects participated; 24 medical students, 12 BST's and 5 expert surgeons. Demographics for the medical student groups in terms of age and gender are outlined in Table 3.3. Demographic details and prior operative exposure for the 12 BST's are displayed in Table 3.4.

Table 3.3 Demographics of medical students in the laparoscopic appendicectomy group

	High Innate Ability (N = 12)	Low Innate Ability (N = 12)	p value (Mann-Whitney Test)
Age (years)			
Range	18-28	18-25	NS
Mean	20.9	20.7	
St. Dev.	2.7	2.2	
Gender (%)			
Male	50	8	0.02
Female	50	92	
Dominant Hand (%)			
Right	100	100	NS
Left	0	0	
Corrected Vision (%)			
Yes	67	50	NS
No	33	50	
Video Games (%)			
Yes (at least one hr/week)	50	42	NS
No	50	58	
Music (%)			
Yes (achieved distinction)	100	58	0.003
No	0	42	
Sport (%)			
Yes (intercollegiate level)	33	42	NS
No	67	58	

Table 3.4 Demographics of BSTs in the laparoscopic appendicectomy group

	BST's (N = 12)
Age (years)	
Range	25-31
Mean	26.5
St. Dev.	1.8
Gender (%)	
Male	50
Female	50
Dominant Hand (%)	
Right	92
Left	8
Corrected Vision (%)	
Yes	42
No	58
Video Games (%)	
Yes (at least one hr/week)	58
No	42
Music (%)	
Yes (achieved distinction)	83
No	17
Sport (%)	
Yes (intercollegiate level)	50
No	50
Operations Performed Unsupervised (Mean Number)	
Excision of Lesion	28
OGD	1.6
Laparoscopic Appendicectomy	0

3.4.2 Aptitude Distribution Among Medical students

There was a significant difference in the mean aptitude scores of group A and group B in all areas of aptitude. For visual spatial aptitude, the mean score was 82% for group A compared to 30% in group B ($p<0.0001$) (Figure 3.1). The mean score for depth perception for group A and B was 90% and 54% respectively ($p=0.028$) (Figure 3.2). For psychomotor ability, group A had a mean score of 57 seconds compared to 71 seconds for group B ($p=0.049$) (Figure 3.3). A faster time implied a higher aptitude for psychomotor ability.

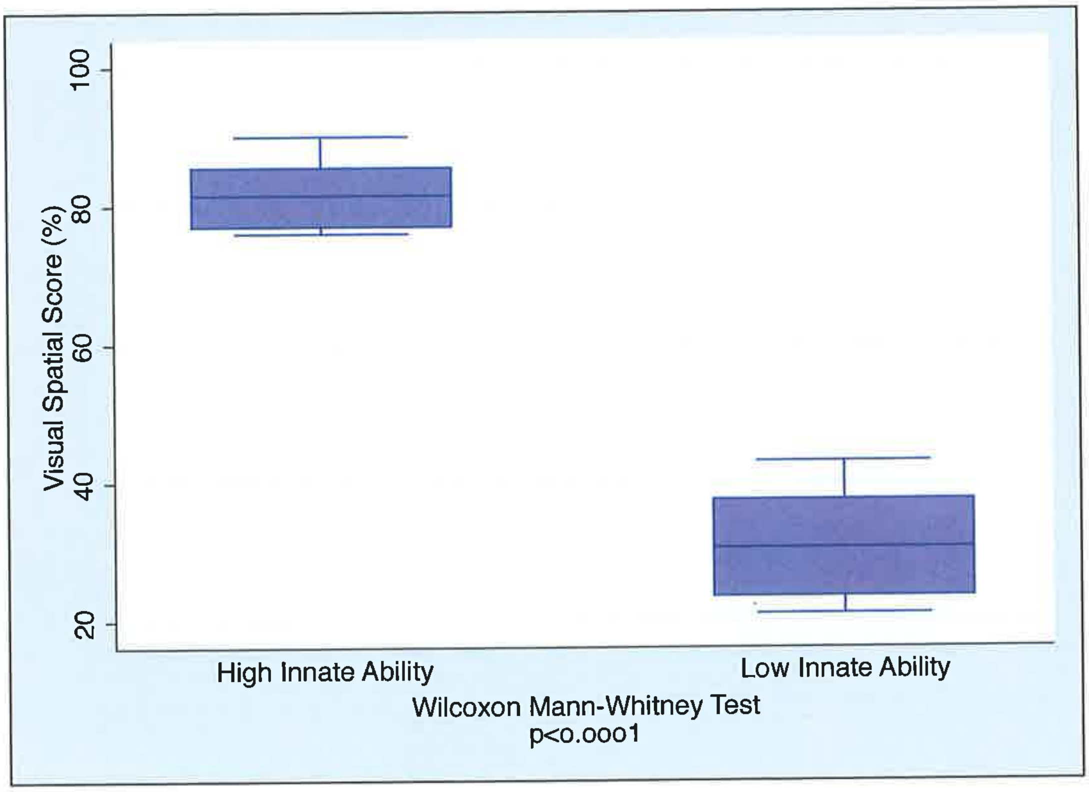


Figure 3.1. Visual spatial scores of medical students in the laparoscopic appendicectomy group

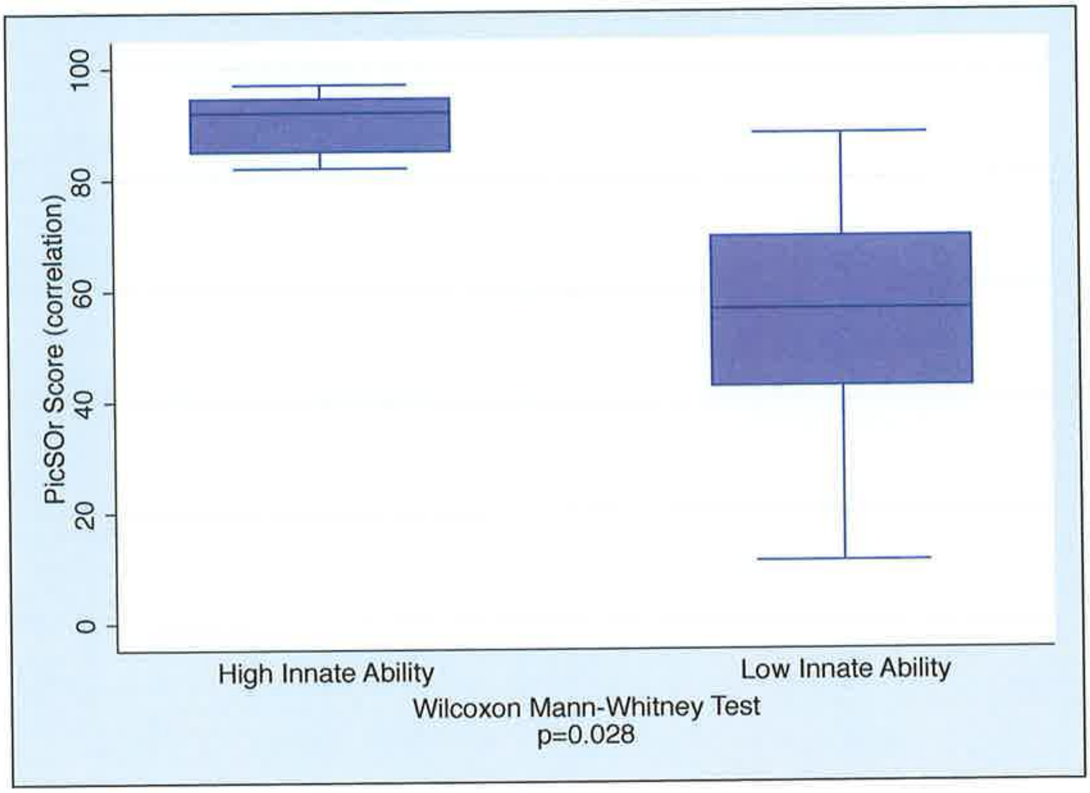


Figure 3.2 Perceptual scores of medical students in the laparoscopic appendicectomy group

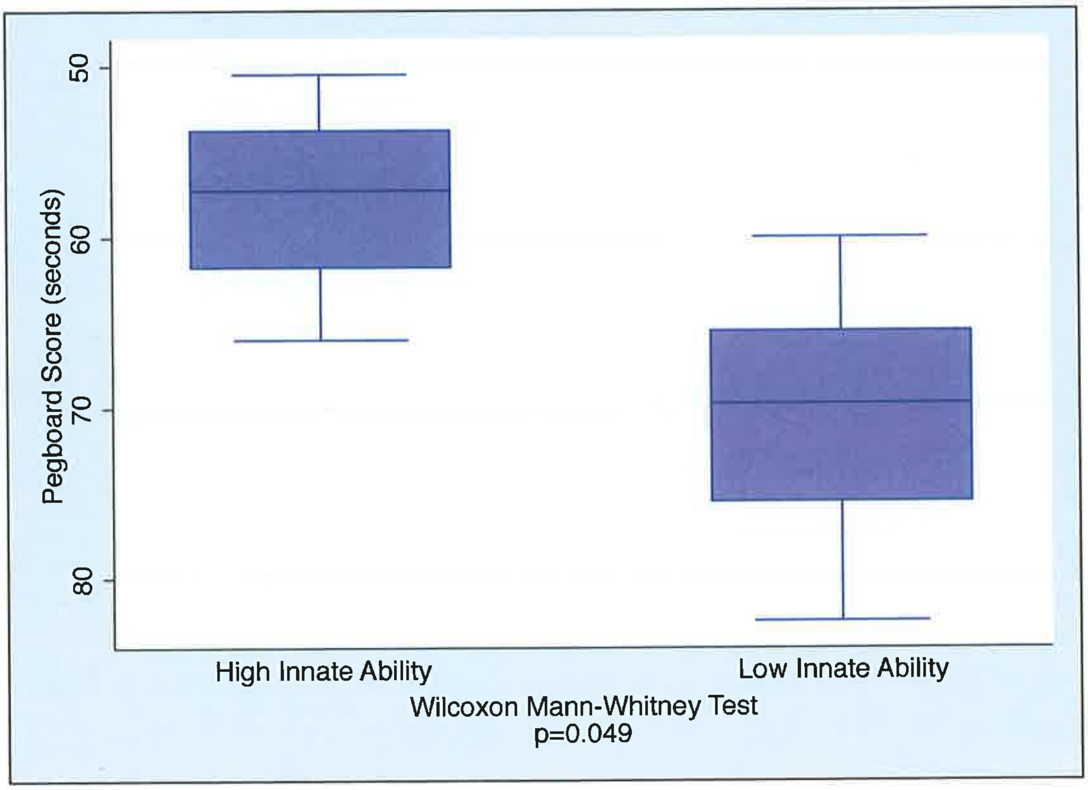


Figure 3.3 Psychomotor scores of medical students in the laparoscopic appendicectomy group

3.4.3 Correlation between Aptitude and Baseline Performance in Medical Students

We found that those with a high aptitude had superior baseline scores for path length (p=0.014), error score (0.012) and OSATS score (<0.0001). However there was no statistically significant difference in the baseline scores for time and smoothness. The metric scores of both groups for their initial attempt are shown in Table 3.5.

Table 3.5 Baseline assessment scores of medical students in laparoscopic appendicectomy

Attempt 1	High Innate Ability (N = 12)	Low Innate Ability (N = 12)	p value (Mann-Whitney Test)
Time, s	1126	1223	0.4
Path Length, mm	33419	57034	0.014
Smoothness	5518	6124	0.37
Error Score (out of 6)	2.2	3.8	0.012
OSATS Score (out of 35)	17	8	<0.0001

3.4.4 Correlation between Aptitude and Attainment of Proficiency in Medical students

There was a significant difference in the number of attempts required to achieve proficiency between the two groups, which is demonstrated in Figure 3.4. Group A achieved proficiency after a mean of 6 attempts. However group B did not achieve proficiency until a mean of 12 attempts ($p<0.0001$). Within group B, three candidates (25%) failed to achieve proficiency after 18 consecutive attempts.

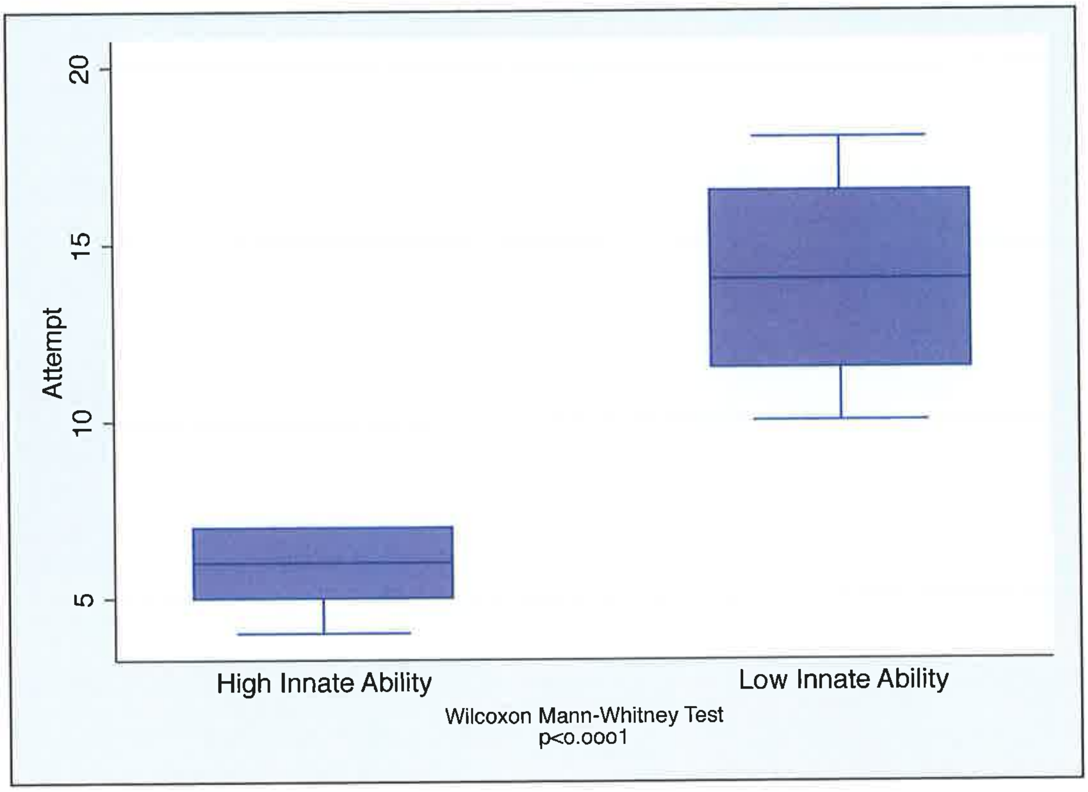


Figure 3.4 The number of attempts required by medical students to achieve proficiency in laparoscopic appendicectomy

3.4.5 Medical Students Path Length Scores

The path length scores for all attempts in both groups are depicted in Figure 3.5. Group A had a higher score at baseline (p=0.014) and they achieved proficiency faster (p<0.0001) than group B. Group B can be divided into two subgroups based on the mean number of attempts, 75% of this group achieved proficiency at a mean of 12 attempts but 25% failed to progress after 18 attempts (Figure 3.6).

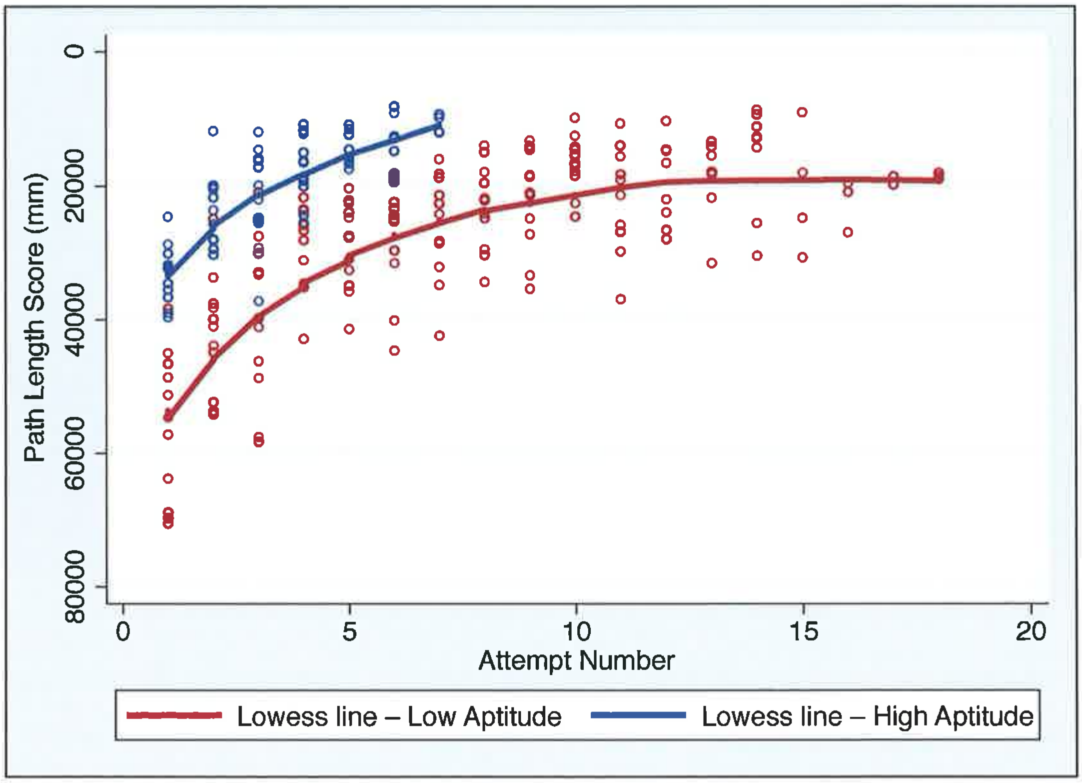


Figure 3.5 Path length scores of medical students in the laparoscopic appendicectomy group

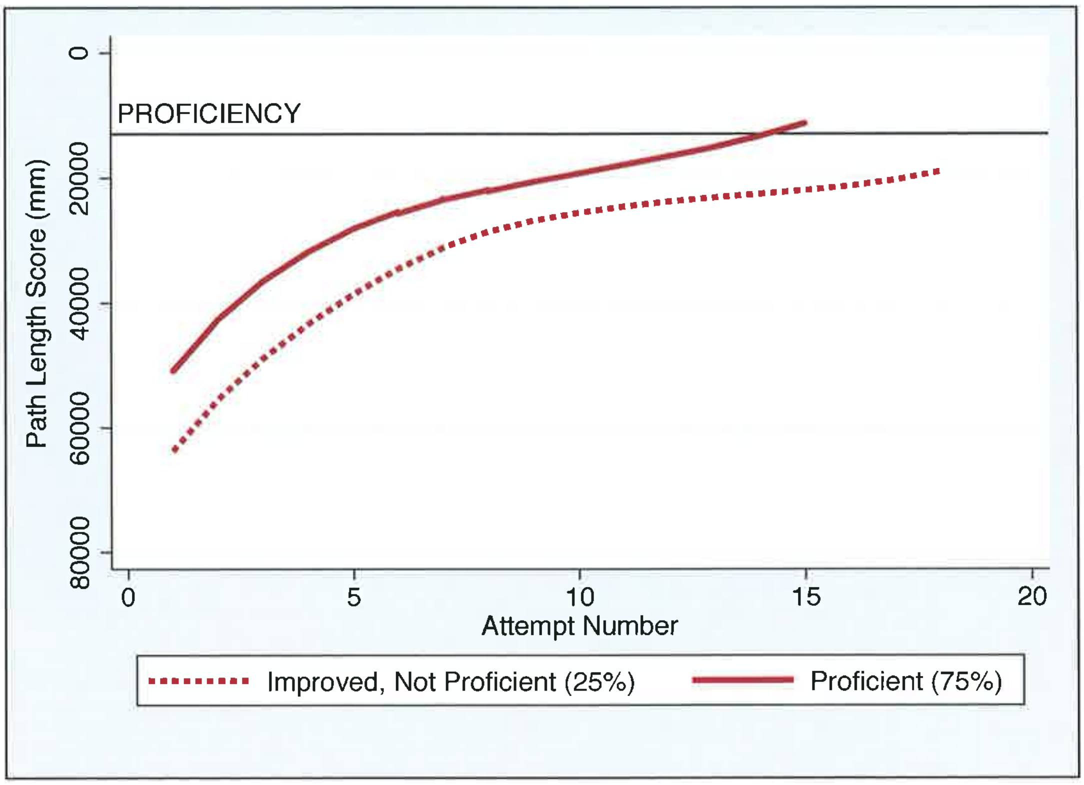


Figure 3.6 Path length scores in medical students with low aptitude in the laparoscopic appendicectomy group

3.4.6 Medical Students Economy of Movement Scores

Economy of movement scores for both groups are displayed in Figure 3.7. Scores for both groups were equal at the initial attempt but group A had a shorter learning curve. Again, group B can be further divided into two learning curves with 75% achieving proficiency and 25% failing to progress, which is shown in Figure 3.8.

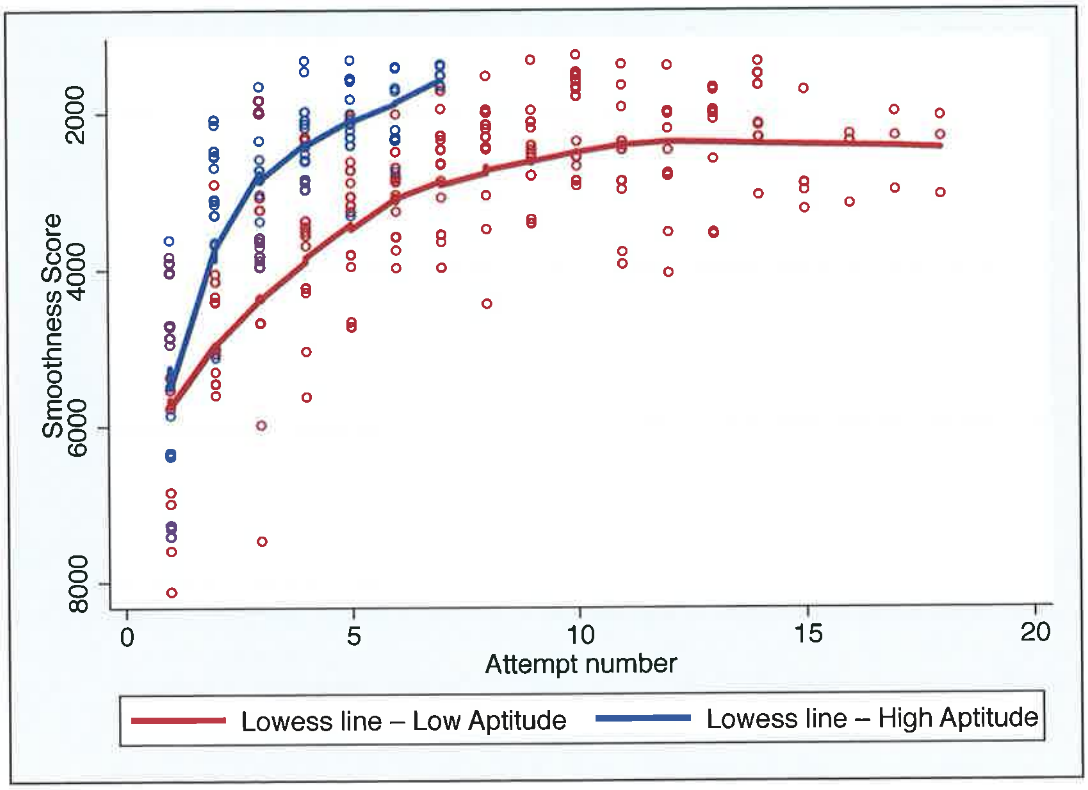


Figure 3.7 Smoothness scores of medical students in the laparoscopic appendicectomy group

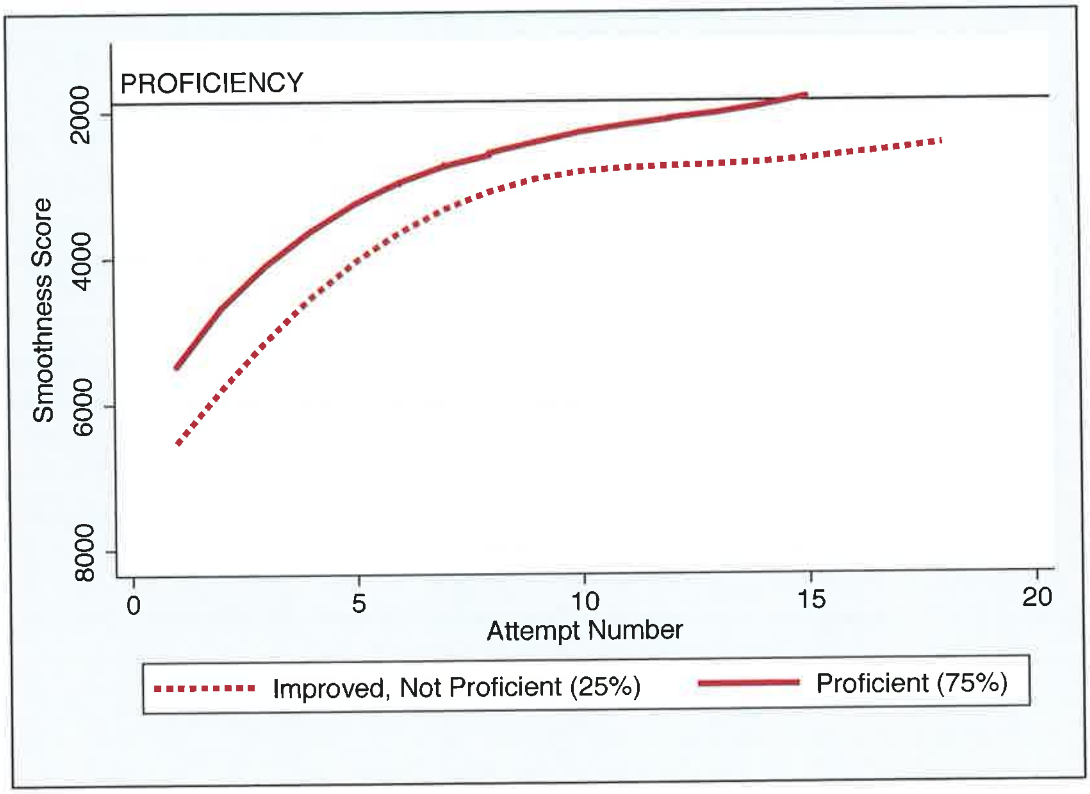


Figure 3.8 Smoothness scores of medical students with low aptitude in the laparoscopic appendicectomy group

3.4.7 Medical Students Time Scores

Time scores for both groups are displayed in Figure 3.9. Scores for both groups were equal at the initial attempt but group A achieve proficient levels in a shorter time frame ($p<0.0001$). Only 75% of group B achieved proficiency as with path length and economy of movement scores and 25% failed to progress, these two separate learning curves are displayed in Figure 3.10.

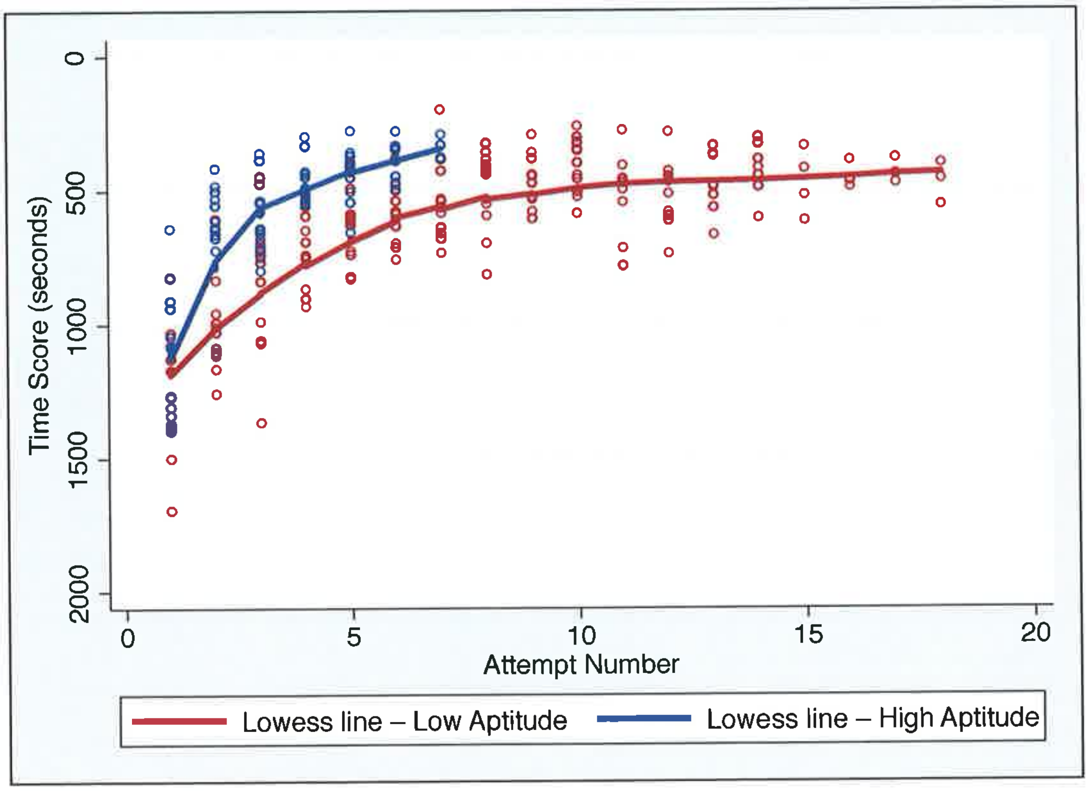


Figure 3.9 Time scores of medical students in the laparoscopic appendicectomy group

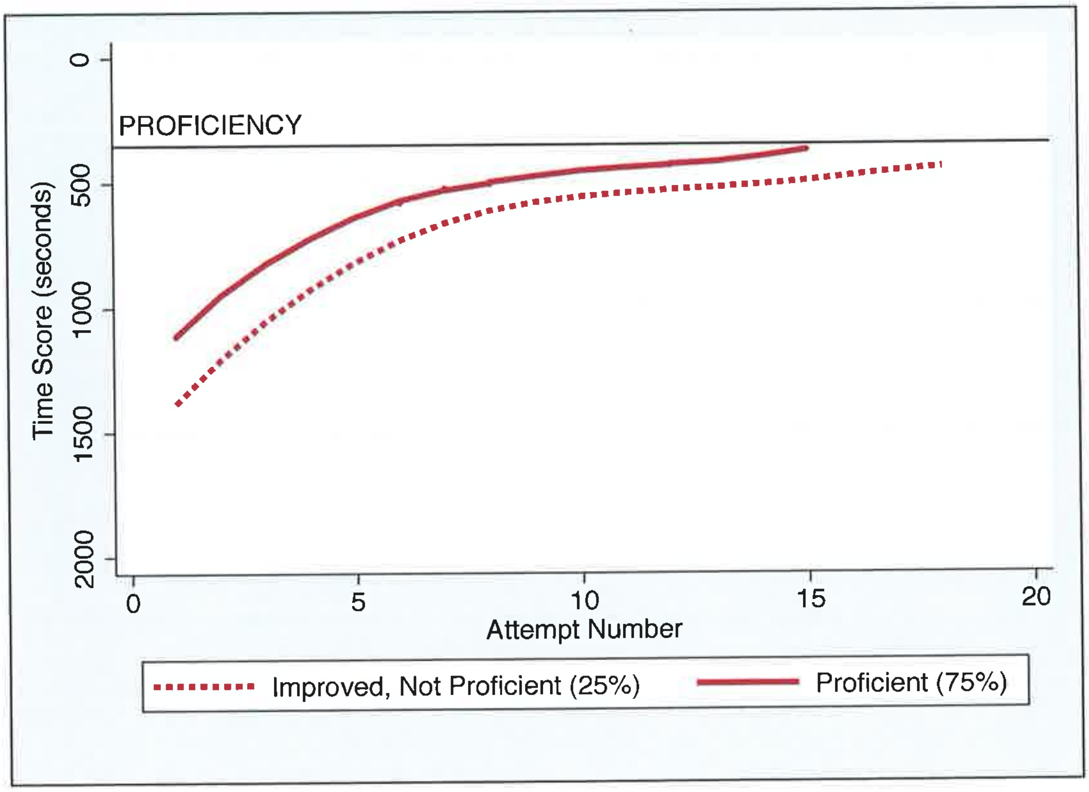


Figure 3.10 Time scores of medical students with low aptitude in the laparoscopic appendicectomy group

3.4.8 Medical Students OSATS Scores

The baseline subjective scores (OSATS, error scores) for both groups are shown in Table 3.5; these are the scores from the first attempt for all candidates. Group A achieved significantly better scores compared to group B in all subjective parameters ($p<0.0001$).

OSATs scores for both groups are shown in Figure 3.11. Group A showed superior performance at baseline ($p<0.0001$) and achieved proficiency with lesser attempts ($p<0.0001$) compared to group B. The consistent separation in learning curves among group B is again present with regard to OSATs scores is depicted in Figure 3.12. Inter-rater reliability between the two assessors was determined using intraclass correlation coefficient (ICC) and was found to be 0.90.

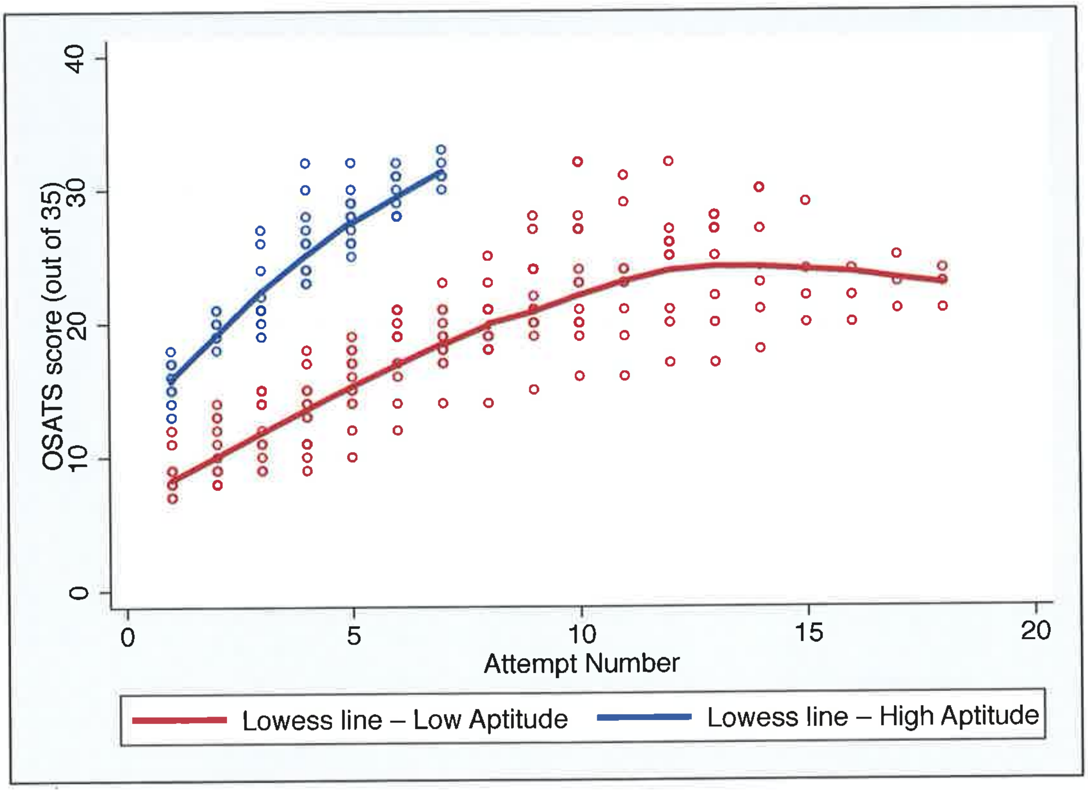


Figure 3.11 OSATS scores of medical students in the laparoscopic appendicectomy group

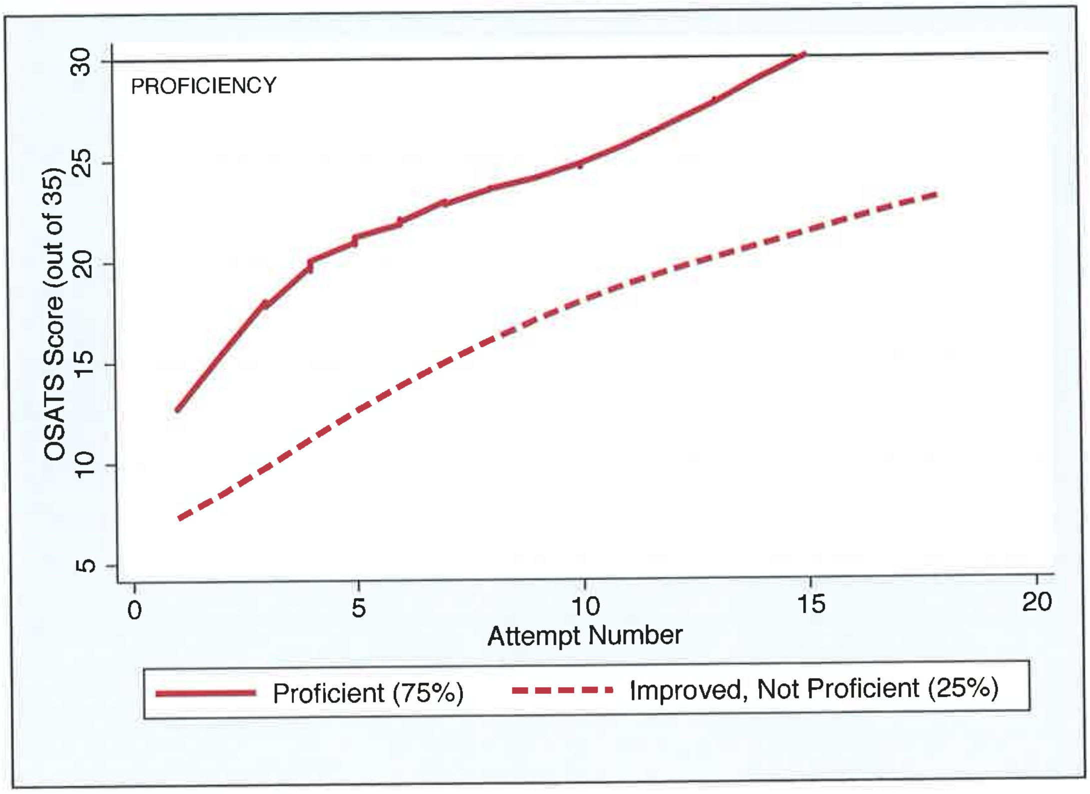


Figure 3.12 OSATS scores of medical students with low aptitude in the laparoscopic appendicectomy group

3.4.9 Medical Students Error Scores

The error scores for both groups are shown in Figure 3.13. Group A committed less errors during their initial attempt (p=0.012) and achieved an error score of 0% at a faster rate than those in group B (p<0.0001). Inter-rater reliability using ICC was 0.95.

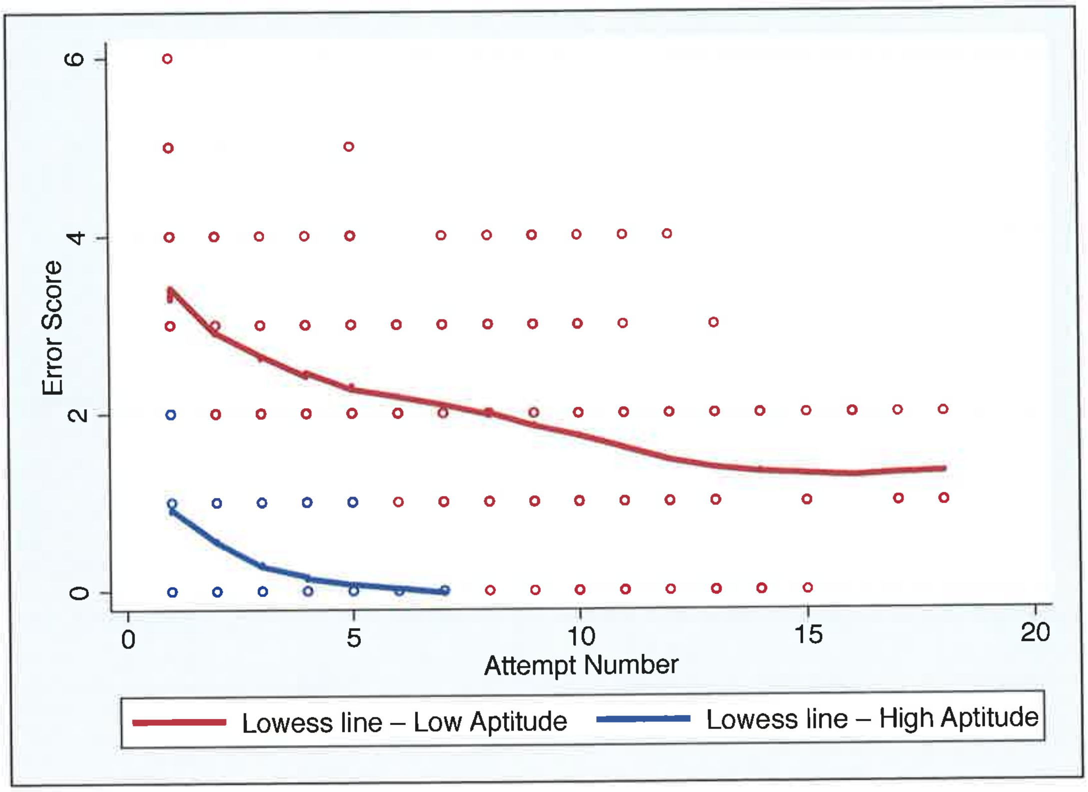


Figure 3.13 Error scores of medical students in the laparoscopic appendicectomy group

3.4.10 Correlation between Aptitude and Baseline Performance in BST's

The results of the correlation between aptitude and baseline performance in BST's is shown in Table 3.6. Again, like attainment of proficiency there was a significant correlation between baseline performance and visual spatial aptitude. However the correlation of baseline performance with depth perception aptitude and psychomotor ability was non significant.

Table 3.6 Correlation between aptitude and baseline performance of BST's in laparoscopic appendicectomy

Aptitude (N = 12)	Baseline Performance	Correlation (r)	p value
Visual-Spatial	Time	$r = 0.5201$	p = 0.083
	Path Length	$r = 0.5819$	p = 0.0472
	Smoothness	$r = 0.7149$	p = 0.009
Depth Perception	Time	$r = 0.0091$	NS
	Path Length	$r = 0.2884$	NS
	Smoothness	$r = 0.2476$	NS
Psychomotor	Time	$r = 0.3283$	NS
	Path Length	$r = 0.2476$	NS
	Smoothness	$r = 0.4267$	NS

3.4.11 Correlation between Aptitude and Attainment of Proficiency in BST's

We assessed the relationship between aptitude and ability to reach the predefined proficiency scores in laparoscopic appendicectomy. There was a significant correlation between the number of attempts required to achieve proficiency in laparoscopic appendicectomy and visual spatial aptitude. These results are shown in Table 3.7

Table 3.7 Correlation between aptitude and attainment of proficiency of BSTs in laparoscopic appendicectomy

Aptitude (N = 12)	Proficiency	Correlation (*r*)	p value
Visual-Spatial	Number of Attempts	$r = 0.6924$	p = 0.0126
Depth Perception	Number of Attempts	$r = 0.5712$	p = 0.0524
Psychomotor	Number of Attempts	$r = 0.0986$	NS

3.4.12 BST's Metric Scores

The surgical trainees path length, economy of movement and time scores are shown in Figures 3.14, 3.15 and 3.16 respectively. They achieved proficiency after an average of 6 attempts (range 4-8).

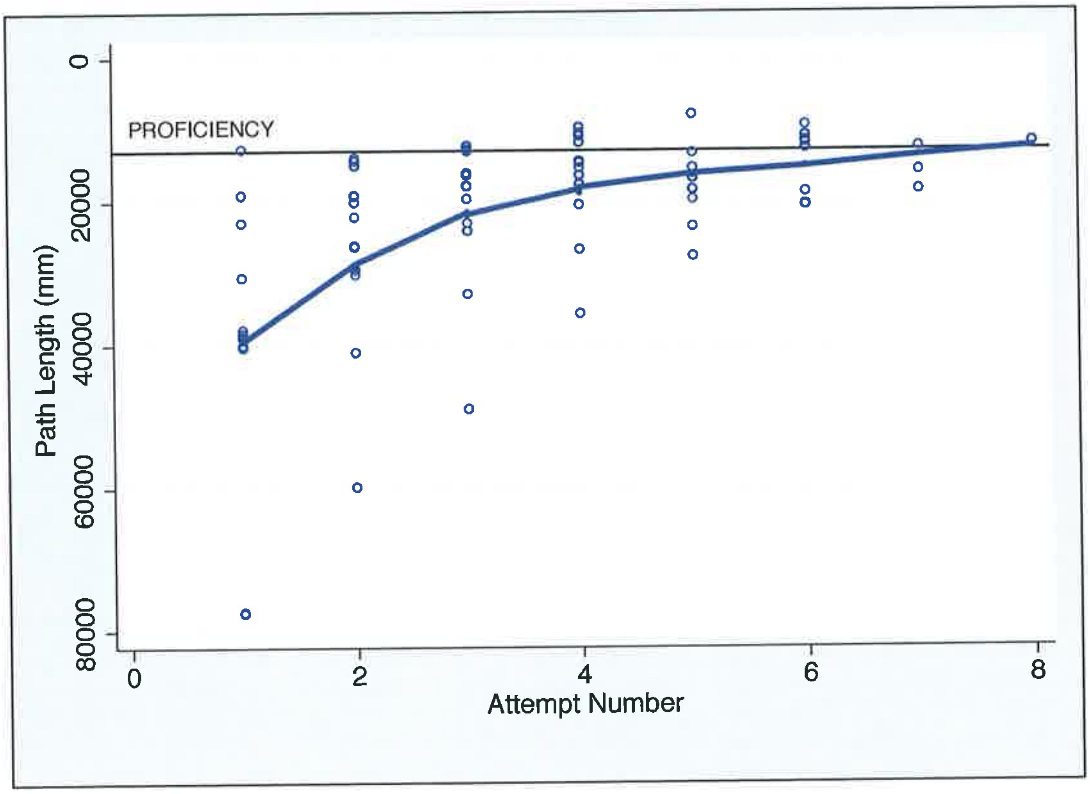


Figure 3.14 BST path length scores in the laparoscopic appendicectomy group

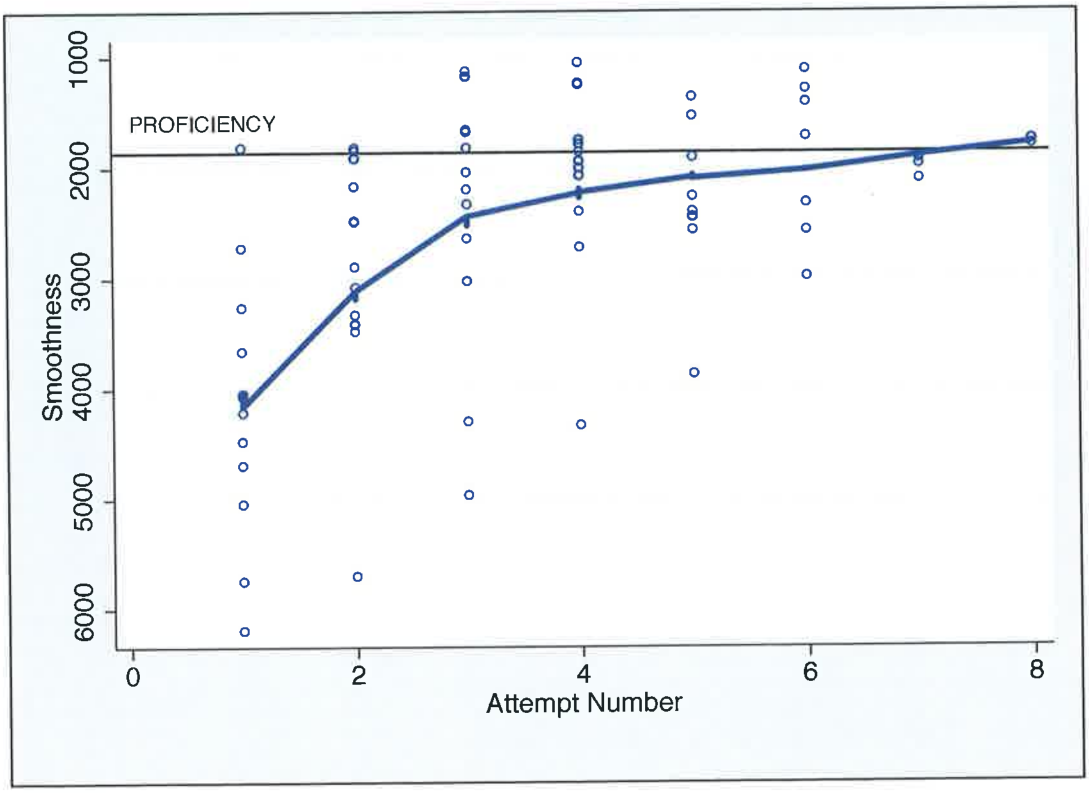


Figure 3.15 BST smoothness scores in the laparoscopic appendicectomy group

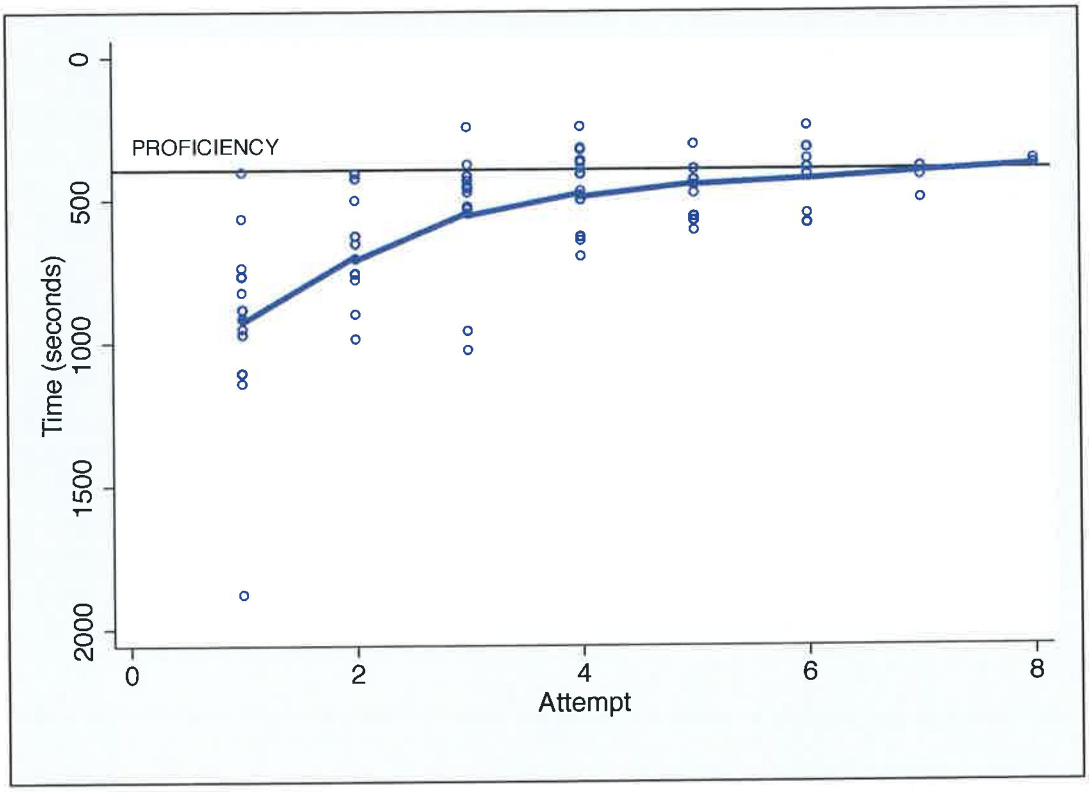


Figure 3.16 BST time scores in the laparoscopic appendicectomy group

3.4.13 BST's Subjective Scores

OSATS scores are illustrated in Figure 3.17, the average baseline score was 15 compared to 33.5 at achievement of proficiency. Inter-rater reliability between the two assessors who rated the laparoscopic appendicectomy procedure was determined using ICC and was found to be 0.92.

The error scores are shown in Figure 3.18. The baseline mean error score was 2.3. Inter-rater reliability using ICC was 0.96.

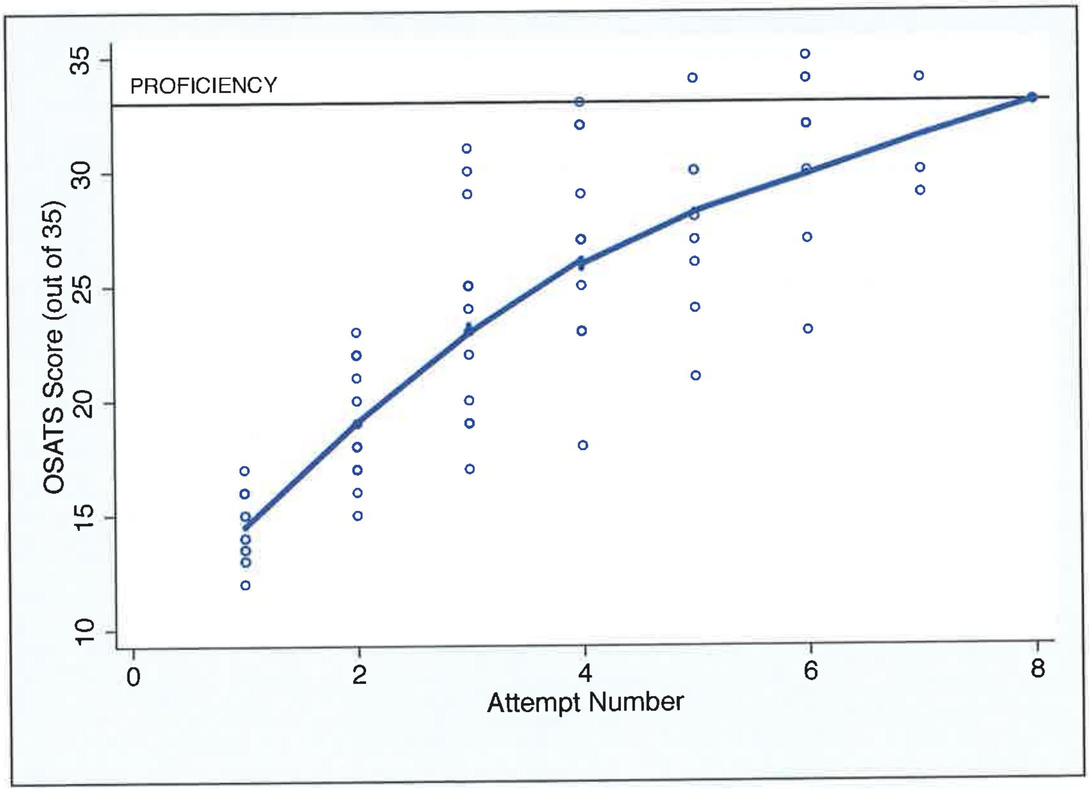


Figure 3.17 BST OSATS scores in the laparoscopic appendicectomy group

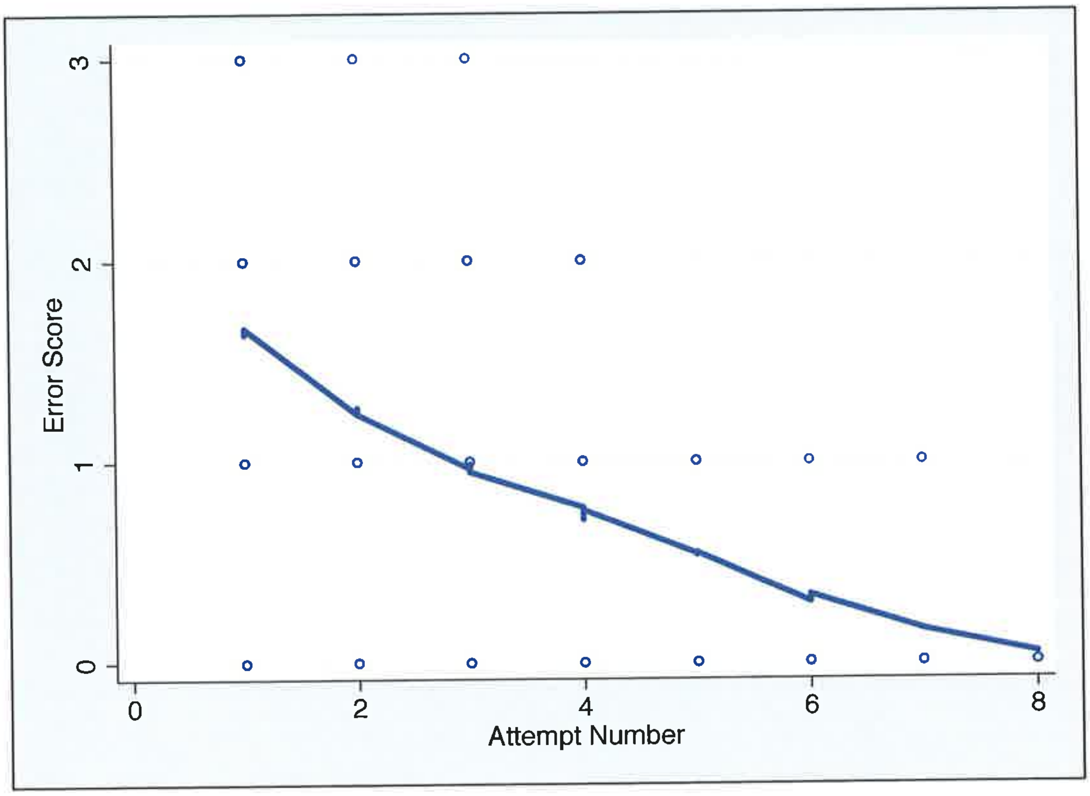


Figure 3.18 BST error scores in the laparoscopic appendicectomy group

3.4.14 Prediction of the Learning Curve for Laparoscopic Appendicectomy

Significant differences were found between the average expert score and the average score of the 12 BST's for the first simulated laparoscopic appendicectomy performed (p=0.009). Scores improved with each subsequent procedure and reached expert scores on an average of 6 (range 4-8) attempts. Significant differences were found for time (p=0.018), path length (p=0.018), smoothness (p=0.018) and error score (p=0.0169) when the mean scores of the first attempt were compared with the sixth attempt. These results are shown in Table 3.8

Table 3.8 Baseline and final assessment scores of BST's in laparoscopic appendicectomy

	Attempt 1	Attempt 6	p value (Wilcoxon signed-rank test)
Time	928	378	0.018
Path Length, mm	39428	11965	0.018
Smoothness	4190	1780	0.018
OSATS Score (out of 35)	15	33.5	0.0001
Error Score (out of 6)	2.3	0	0.0169

3.5 Discussion

In this chapter, we evaluated the impact of innate ability on achieving proficiency in laparoscopic appendicectomy. By directly comparing two groups of medical students with different aptitude scores who otherwise had similar demographics, we were able to eliminate any confounding factors. The results demonstrate that medical students with high innate ability achieve proficiency in laparoscopic appendicectomy twice as fast as those with low innate ability. This was reproducible across all measurable parameters including path length, smoothness, time, OSATS and error scores.

Training a surgeon involves considerable time commitments and has significant financial implications. The aim of a surgical training program is to provide a medium to train candidates efficiently, while ensuring that they are proficient in their chosen field at completion of training. As we strive to improve surgical training to ensure minimal attrition rates and achieve technical proficiency upon completion of the program in a time efficient manner, we should select those with the appropriate aptitude. The results of the current study confirm that subjects with high aptitude achieve proficiency significantly earlier than those with inferior aptitude. Not all candidates with low aptitude can achieve proficiency despite multiple attempts. One in four candidates in the group with low aptitude were unable to achieve proficiency in laparoscopic appendectomy despite 18 attempts of the procedure. These findings have significant resource implications.

Work previously carried out by Stefanidis et al (Stefanidis et al., 2006) concluded that the importance of psychomotor testing lies within the prediction of how rapidly one can acquire a laparoscopic skill. Our work specifically aimed at determining the exact difference in rate of skill acquisition between those with contrasting aptitude. We have demonstrated that previously validated aptitude tests can be utilized to predict those who will demonstrate the technical ability and learning facility to achieve proficiency from a 'novice phase'. Current literature shows that residents with higher visual-spatial scores perform significantly better than did those with lower scores (Wanzel et al., 2002, Wanzel et al., 2003). Our study strongly supports these findings as we have shown that those with high aptitude scores had a reduced error score, superior baseline path length and quicker attainment of proficiency. Our findings also support previous work that demonstrated that not all surgical candidates could achieve proficiency. Grantcharov's study (Grantcharov and Funch-Jensen, 2009) found that 8.1% of his population group was unable to learn the laparoscopic technique.

We also aimed to map the number of attempts required by trainee surgeons to reach technical proficiency in laparoscopic appendicectomy using simulation-based training. An average of 6 attempts was required to reach benchmark goals and all objective metrics improved from baseline to proficiency. The attainment of proficiency and baseline performance was correlated with the three areas of aptitude for the surgical trainees. Visual spatial ability was found to significantly correlate with the number of attempts to attain proficiency and a superior baseline performance. However depth perception and psychomotor ability did not correlate significantly.

The learning curve for laparoscopic appendicectomy was found to be 6 attempts for both the surgical trainees and the medical students with high aptitude. None of the surgical trainees had similar scores to the low innate ability group. Therefore it could be possible that the trainees are perhaps self-selecting themselves into a career in surgery based on innate ability.

A simulation based proficiency based progression programme is core in our surgical training pathway. These findings suggest that the resource allocation for proficiency based technical training in surgery may need to be tailored according to a trainees natural ability.

In conclusion, candidates with high innate ability became proficient at completing a laparoscopic appendectomy at a faster rate than those with lesser innate ability. Our data provides compelling support for an objective multifaceted selection process to select suitable trainees for future training programs.

Chapter 4

Impact of Innate Ability on Completion of the Learning Curve for a Complex Laparoscopic Task

4.1 Introduction

The technique of laparoscopic suturing has been well described in the literature by Szabo and colleagues (Szabo et al., 1994). Laparoscopic suturing is an advanced skill that enables the surgeon to broaden the application of laparoscopy (Allen et al., 2003). However, the skill is difficult to acquire and requires specialized training (Stefanidis et al., 2010). One of the difficulties encountered when learning such a complex task such as laparoscopic suturing and intra-corporeal knot tying is that it requires a very high level of technical ability; which can be better developed in a simulated model (Korndorffer et al., 2005) .

Previous studies have shown that the more complex the procedure, the steeper the associated learning curve (Tekkis et al., 2005). In particular a protracted learning curve has been demonstrated in advanced laparoscopic techniques, such as laparoscopic suturing (Van Bruwaene et al., 2009).

Botden et al examined the learning curve for laparoscopic suturing (Botden et al., 2009b), they found that the number of repetitions required to reach the top of the performance curve was eight. We planned to establish if this learning curve differed depending on the subject's aptitude.

Several studies have examined the relationship between specific areas of aptitude such as visual-spatial and perceptual ability with laparoscopic technical skill performance (Gallagher et al., 2003, Hassan et al., 2007). These studies have concluded that superior laparoscopic performance is demonstrated among novice surgical trainees who possess such attributes. In 2009 Grantcharov et al (Grantcharov and Funch-Jensen, 2009) assessed the learning curve patterns in acquiring laparoscopic skills. This study concluded that the familiarization rate of laparoscopic techniques varies according to psychomotor ability and four types of learning curves were identified.

We have previously shown in Chapter 3 that aptitude can predict the rate at which a surgical novice achieves proficiency in a basic laparoscopic task such laparoscopic appendicectomy. Intuitively, the more complex a laparoscopic procedure/task is, than the greater the impact that innate ability will have on achieving proficiency.

Laparoscopic suturing and intracorporeal knot tying were chosen for assessment. Although they are not performed in a wide range of laparoscopic procedures, they are easily comprehensible yet are some of the most technically demanding laparoscopic tasks therefore ideal for teaching in an experimental setting.

We aimed to compare the rate at which two groups of medical students became proficient in a laparoscopic suturing and intra-corporeal knot-tying task. These two groups were at opposite ends of the aptitude spectrum.

4.2 Objectives

4.2.1 Hypothesis

We have shown in Chapter 3 that aptitude affects the rate at which one becomes proficient in a basic laparoscopic task. Following on from this, we wanted to ascertain the degree to which aptitude played a role in achieving proficiency in an advanced complex laparoscopic task.

It was established in Chapter 3 that increasing aptitude predicts superior baseline performance in path length performance scores as well as a reduced error score for laparoscopic appendicectomy. We hoped to investigate if this also applied to a more complex MIS task or if there was an affect on baseline performance across all parameters.

Three candidates were unable to achieve proficiency despite repeated practice in laparoscopic appendicectomy. We aimed to assess if there was a greater rate of failure to achieve proficiency in a more complex MIS task.

Trainees are afforded limited learning opportunities in the operative setting due to a complex array of challenges that our healthcare system currently faces. They are afforded even less opportunity to perform complex tasks and procedures. We hypothesised that a trainee could complete the learning curve for laparospopic suturing using a simulation based training programme. Our aim was to establish the numbers of attempts required to reach pre-established expert goals.

4.2.2 Detailed Objectives

Objective 1. To investigate the impact that innate ability has on completion of the learning curve for laparoscopic suturing.

We aimed to compare the rate at which two groups of medical students became proficient in a laparoscopic suturing and intracorporeal knot-tying task. These two groups were at opposite ends of the aptitude spectrum.

Objective 2. To determine if medical students with high innate ability have superior baseline performance in laparoscopic suturing than those with low ability.

In the same two groups of medical students, we aimed to compare metric, error, OSATS and FLS scores of the first attempted laparoscopic suturing task performed between the two groups.

Objective 3. To map the number of attempts required by trainee surgeons to reach technical proficiency in laparoscopic suturing.

We aimed to evaluate the learning pathway of surgical trainees by establishing the number of attempts required to reach predefined goals. By doing this we aimed to determine the minimum number of practice attempts, which must be performed by trainees during their training.

Objective 4. To set benchmark proficiency levels for laparoscopic suturing on the ProMIS simulator.

In order to map the learning curve for laparoscopic suturing, we needed to establish proficiency levels. We aimed to recruit a group of expert consultant surgeons, who had performed greater than 150 complex laparoscopic procedures.

4.3 Materials and Methods

4.3.1 Participant Recruitment

As outlined in chapter 2, participants were recruited to take part in this study on a voluntary basis. An email was distributed to all medical students in RCSI with details of the research project. They were given the opportunity to undergo psychometric testing in order to select candidates for the reserch study. The fifty-two students who replied were asked to complete four visual spatial tests (card rotation, cube comparison, map planning and spatial orienation), the grooved pegboard test and the picsor test which took place in RCSI, 121 St. Stephen's Green, Dublin 2.

An email was also send out to the first and second year BST's calling for voulunteers to take part in this study. As laparoscopic suturing is considered one of the most complex laparoscopic tasks, both first and second years were elligible for this experiment. However, if any of the trainees had performed laparscopic suturing as part a laparoscopic procedure, they were excluded from the study.

Five laparoscopic experts was also recruited in order to set benchmark proficiency levels on the ProMIS simulator. Each of these experts had performed over 150 complex laparoscopic procedures.

All participants who were selected were asked to sign a consent form allowing all data collected to be used for research purposes. It was made clear to all the subjects that the data was stored and presented in an anonymous format.

4.3.2 Participant Demographics

Based on the results of the aptitude tests administered, twenty students were recruited to participate in this research study. The first group of ten students were considered to have high aptitude (group A) as their score was one standard deviation higher than the average score of all the students tested. The second ten students selected were considered to have low aptitude (group B) as their score was one standard deviation lower than the average score of all students tested. Thus the two groups selected were two standard deviations apart. Ten basic surgical trainees were recruited to take part in the study. The inclusion and exclusion criteria for the study particpants have been previously outlined in chapter 2.

4.3.3 Aptitude Assessment

The three main areas of aptitude considered to be relevant for minimally invasive procedures are visual spatial aptitude, psychomotor aptitude and depth perception ability.

Visual spatial aptitude was assessed using tests sourced from the Kit of Factor Referenced Cognitive Tests (Ekstrom et al, 1976). Four paper based tests were used; card rotations and cube comparison tests assessed spatial orientation, map planning test assesed spatial scanning and surface development test assessed spatial visualisation. Psychomotor aptitude was tested using the Grooved Pegboard which assess manual dexterity and hand-eye coordination (Dikmen at all, 1999). Depth perception was assessed using a computer based software program known as Pictorial Surface Orientation (PicSOr) which tests ones ability to convert a 2D image on a screen to a 3D image (Cowie, 1998). A more detailed explanation of these aptitude assessments and how they were carried out are outlined in chapter 2.

4.3.4 Setting Proficiency Levels

Five laparoscopic experts were recruited to set proficency levels. Each of them had performed over 150 complex laparoscopic procedures. Each expert was asked to perform a laparoscopic suturing task on the ProMIS simulator using the same module and materials and under the same conditions as the medical students. They performed the procedure twice and a mean score of the objective parameters path length, smoothness and time was calculated for each expert. Proficiency was then determined by calculating the mean expert score (Appendix V).

4.3.5 Surgical Performance Assessment

The ProMIS III® Simulator was used for assessment. Further details can be found in Chapter 2.

A 10 x 12 cm piece of synthetic suturing skin (The Chamberlain Group, Massachusetts®) was placed in the simulator tray. The laparoscopic suturing task was performed according to Fundamentals of Laparoscopic Surgery (FLS) principles (Appendix IV) using 3/0 silk and two laparoscopic needle holders. Three knots were thrown in order to complete the task.

Prior to performance assessment each subject received didactic teaching. Each candidate was sent a stepwise approach detailing how to perform the task (Table 4.1) and a video-link to a live recording of a specific part of a laparoscopic rectopexy before the experiment commenced. When the subject attended the first session, a simulated laparoscopic suturing demonstration was conducted. They were allocated time to ask questions before they attempted a mandatory multiple-choice questionnaire to ensure complete comprehension of the procedure. Prior to their first assessement, they had an opportunity to familiarise themselves with the testing equipment by completing a basic laparoscopic task.

Participants were asked to perform the suturing task consecutively until they reached the proficiency scores. An interval practice curriculum rather than massed practice was implemented (Gallagher et al., 2005, Metalis, 1985). Subjects were allowed to perform a maximum of four tasks per session to avoid fatigue (Van Dongen et al., 2011b, Verdaasdonk et al., 2007). Sessions were carried out at a maximum interval of two weeks (Stefanidis et al., 2008). Subjects were supervised by a senior surgeon at all times and if the subject needed guidance, instructions were given. The senior surgeon did not take over the task at any point during the performance assessment. Upon completion of each procedure, the simulator provided a summary score report which was relayed to the subject so that they were aware of their progress throughout the experiment.

Each performance was recorded and subsequently assessed by two reviewers, blinded to the status of the surgical novice, using the Objective Structured Assessment of Technical Skill (OSATS) scoring system (Faulkner et al., 1996) and the FLS rating scale. The quality of each knot was assessed after task completion. This information was relayed to the candidate after each performance.

Each tray was examined after procedure completion for four pre-defined errors (Table 4.2) and this was also relayed to the subject after each performance.

4.3.6 Attainment of Proficiency

The 30 subjects consecutively performed the task until proficiency was reached. Simulator calculated metrics on the ProMIS simulator along with error, OSATS and FLS scores were recorded for each attempt.

4.3.7 Statistical Analysis

Statistical analysis was performed using Stata 12.0. The aptitude, metric and proficiency scores of the two groups were compared using the Mann Whitney test. A p value <0.05 was considered statistically significant. Non-parametric testing was performed as none of the data was shown to have a normal distribution using the Shapiro Wilk test. Inter-rater reliability was determined using intraclass correlation coefficient (ICC).

Table 4.1. Operative steps of laparoscopic suturing task

1.	**Needle Position**	Needle is held ½ to 2/3 from the tip at a 90 degree angle. One must use the other instrument for stability
2.	**Needle Driving**	Needle is entered to the tissue plane at 60 -90 degrees. You must drive with one movement
3.	**Needle Position 2**	Remove the needle with your second instrument once it has entered through the tissue. Needle is held ½ to 2/3 from the tip at a 90 degree angle.
4.	**Needle Driving 2**	Needle is entered to the tissue plane at 60 -90 degrees. You must drive with one movement
5.	**Pull the Suture**	Pull the suture through the tissue with the needle on the needle holder at all times, stabilising the tissue with the second needle holder
6.	**Knot Tying**	Perform a double throw (c shape) and bring the instruments to opposite ends in order to tighten the knot. Then perform a reverse c loop and and again bring the instruments to opposite ends in order to tighten the knot in a square shape. Then perform one last c loop and bring the instruments to opposite ends in order to tighten the knot in a square shape.

Table 4.2. Laparoscopic suturing task errors

1.	Pulling the suture through the skin prior to knot tying
2.	Tears in the skin
3.	Cutting the suture too long or too short
4.	Slip Knot
5.	Suture too loose or too tight

4.4 Results

4.4.1 Participant Demographics

35 subjects took part; 20 medical students, 10 BST's and 5 expert surgeons. Demographics for the 20 medical students in terms of age, gender, handedness and video game, sporting and musical ability are outlined in Table 4.3. Prior operative experience and demographic details for the 10 BSTs is displayed in Table 4.4.

Table 4.3. Demographics of medical students in the laparoscopic suturing group

	High Aptitude Group N = 10	Low Aptitude Group N = 10	p value (Mann-Whitney Test)
Age (years)			
Range	19-27	19-36	NS
Mean	21.1	23.5	
St. Dev.	2.6	5.5	
Gender (%)			
Male	50	20	NS
Female	50	80	
Dominant Hand (%)			
Right	100	100	NS
Left	0	0	
Corrected Vision (%)			
Yes	70	50	NS
No	30	50	
Video Games (%)			
Yes (at least one hour/week)	90	10	0.001
No	10	90	
Music (%)			
Yes (achieved distinction)	70	50	NS
No	30	50	
Sport (%)			
Yes (intercollegiate level)	80	40	NS
No	20	60	

Table 4.4 Demographics of BSTs in the laparoscopic suturing group

	BST's (N = 10)
Age (years)	
Range	25-30
Mean	26.9
St. Dev.	1.7
Gender (%)	
Male	50
Female	50
Dominant Hand (%)	
Right	90
Left	10
Corrected Vision (%)	
Yes	70
No	30
Video Games (%)	
Yes (at least one hr/week)	60
No	40
Music (%)	
Yes (achieved distinction)	60
No	40
Sport (%)	
Yes (intercollegiate level)	70
No	30
Operations Performed Unsupervised (Mean Number)	
Excision of Lesion	34
Laparoscopic Appendicectomy	1
Laparoscopic Cholecystectomy	0

4.4.2 Aptitude Distribution Among Medical Students

The high aptitude group (group A) achieved significantly higher scores than the low aptitude group (group B) in all three areas of aptitude tested as illustrated in Figures 4.1, 4.2 and 4.3. Group A scored 72% for visual spatial ability compared with 24% in group B ($p<0.0001$). Depth perception scores were 93% in group A compared with 50% in group B ($p<0.0001$) while group A demonstrated improved psychomotor ability by performing the pegboard task in 58 seconds compared with 73 seconds in group B ($p=0.03$).

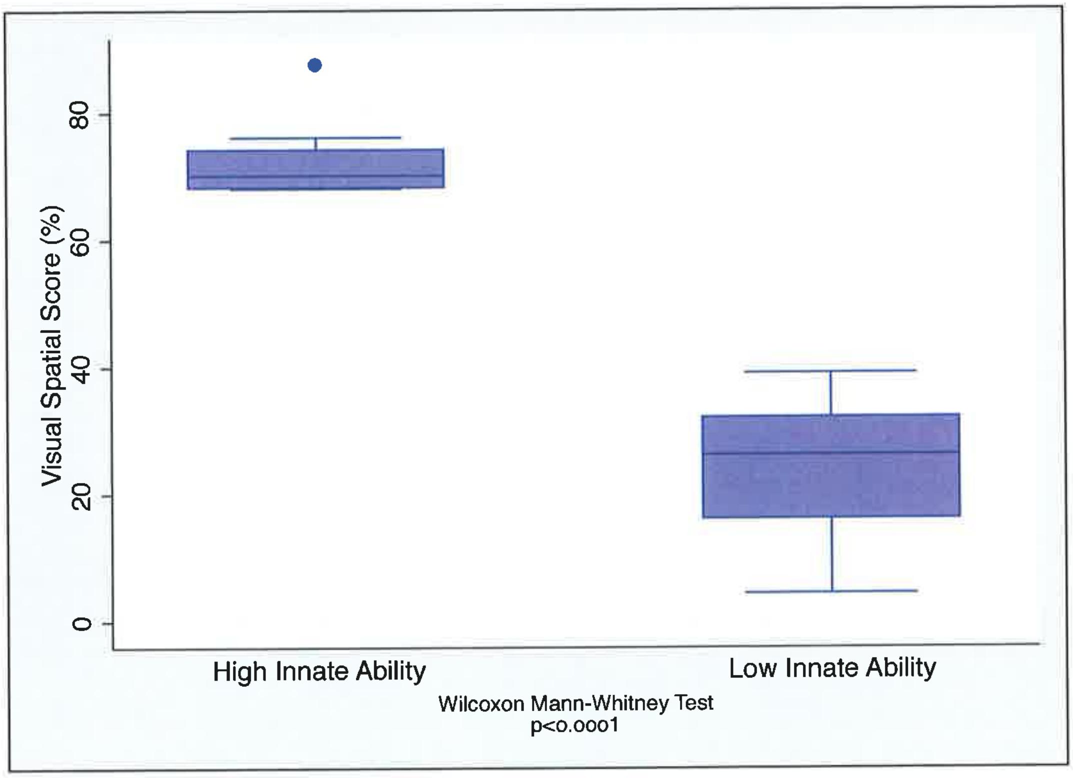


Figure 4.1 Visual spatial scores of medical students in the laparoscopic suturing group

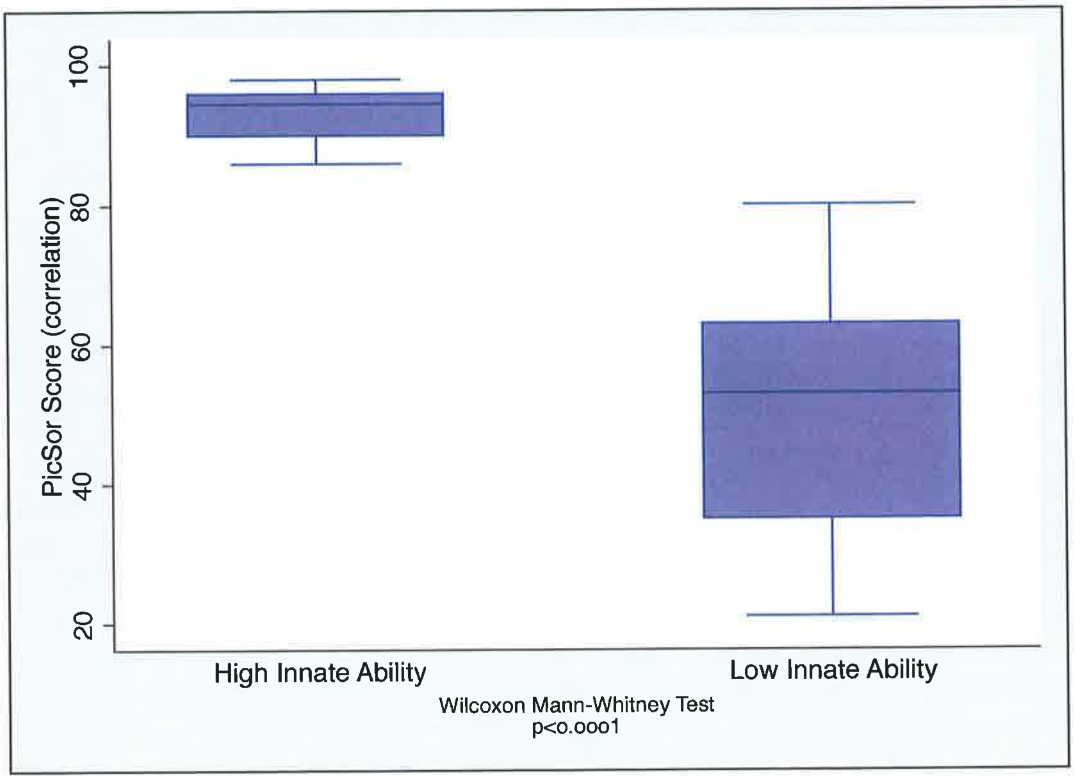


Figure 4.2. Perceptual scores of medical students in the laparoscopic suturing group

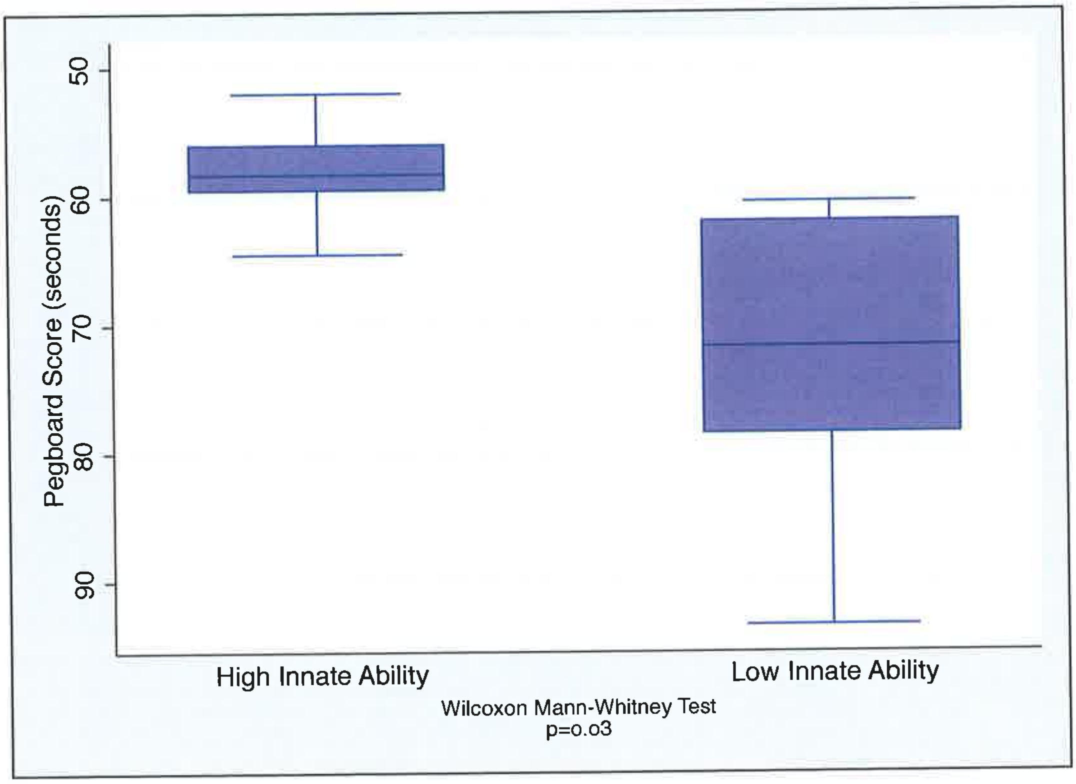


Figure 4.3 Psychomotor scores of medical students in the laparoscopic suturing group

4.4.3 Correlation between Aptitude and Baseline Performance in Medical Students

The baseline metric scores for both groups are shown in Table 4.5; these are the scores from the first attempt for all subjects. Group A achieved better scores than group B in all parameters ($p<0.0001$).

Table 4.5 Baseline assessment scores of medical students in laparoscopic suturing

Attempt 1	High Innate Ability (N = 10)	Low Innate Ability (N = 10)	p value (Mann-Whitney Test)
Time, s	551	1475	<0.0001
Path Length, mm	16635	57542	<0.0001
Smoothness	-1	4978	<0.0001
OSATS Score (out of 25)	12	7	<0.0001
FLS Score (out of 30)	13	9	0.0006
Error Score (out of 5)	2.2	4.7	<0.0001

4.4.4 Correlation between Aptitude and Attainment of Proficiency in Medical Students

The mean number of attempts required to reach proficiency in each group is shown in Figure 4.4. The mean number of attempts to complete the procedure for group A compared with group B was 7 (range 4-10) versus 12 (range 10-15) p=0.01. In group B only 30% achieved proficiency at a mean of 12 attempts, 40% demonstrated improvement but did not attain proficiency and 30% failed to progress and dropped out of the study as they were unable to complete the task after an average of 5 attempts.

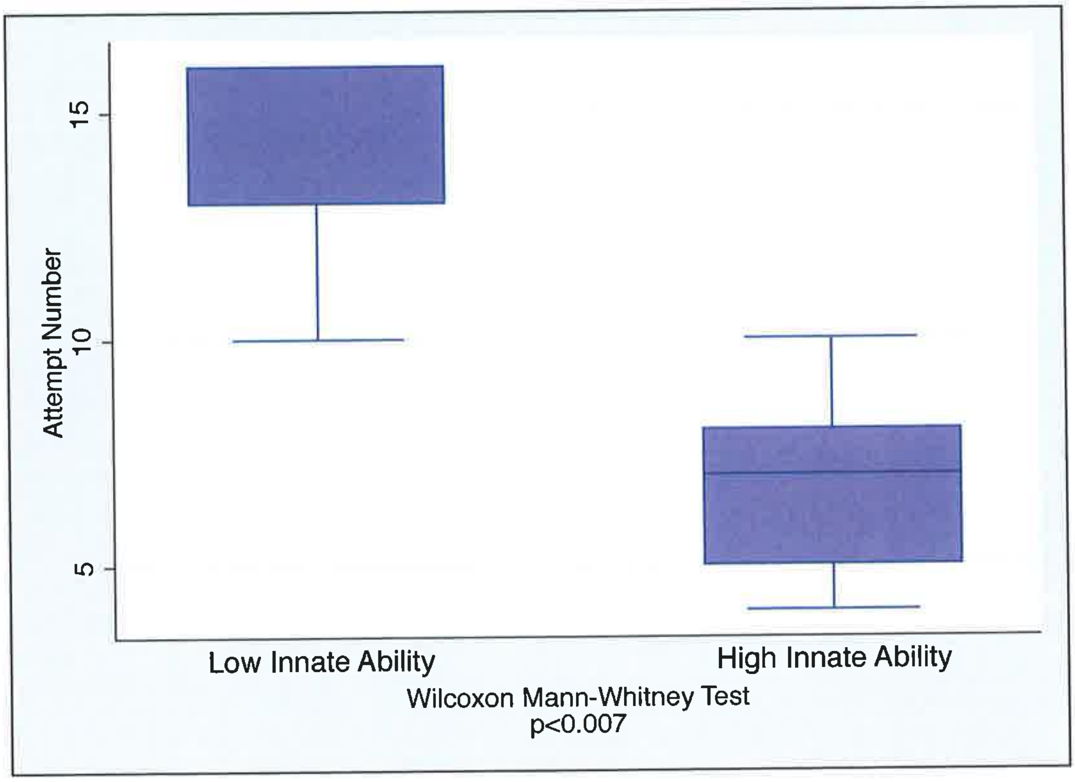


Figure 4.4 The number of attempts required by medical students to achieve proficiency in laparoscopic suturing

4.4.5 Medical Students Path Length Scores

Path length scores are illustrated in Figure 4.4. Group A achieved proficiency faster ($p<0.0001$) than group B. Within group B, 30% achieved proficiency at a mean of 14 attempts while 40% demonstrated improvement but did not attain proficiency despite 16 attempts. 30% failed to progress and were unable to progress along the learning curve (Figure 4.5).

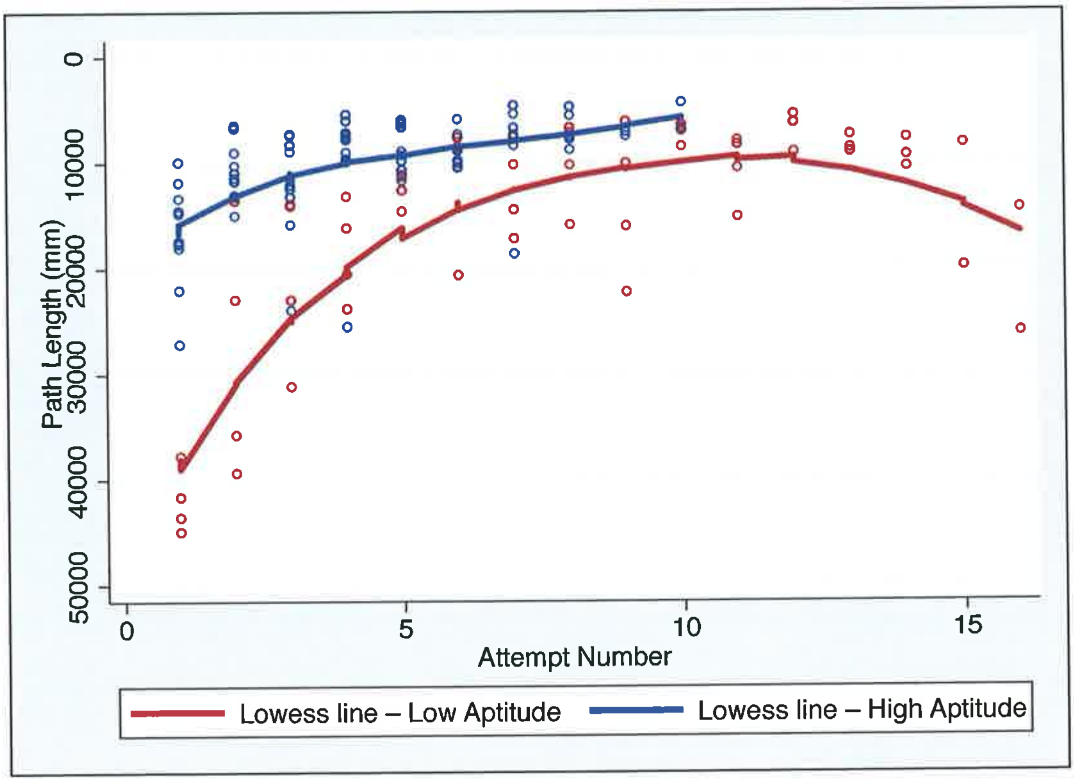


Figure 4.5 Path length scores of medical students in the laparoscopic suturing group

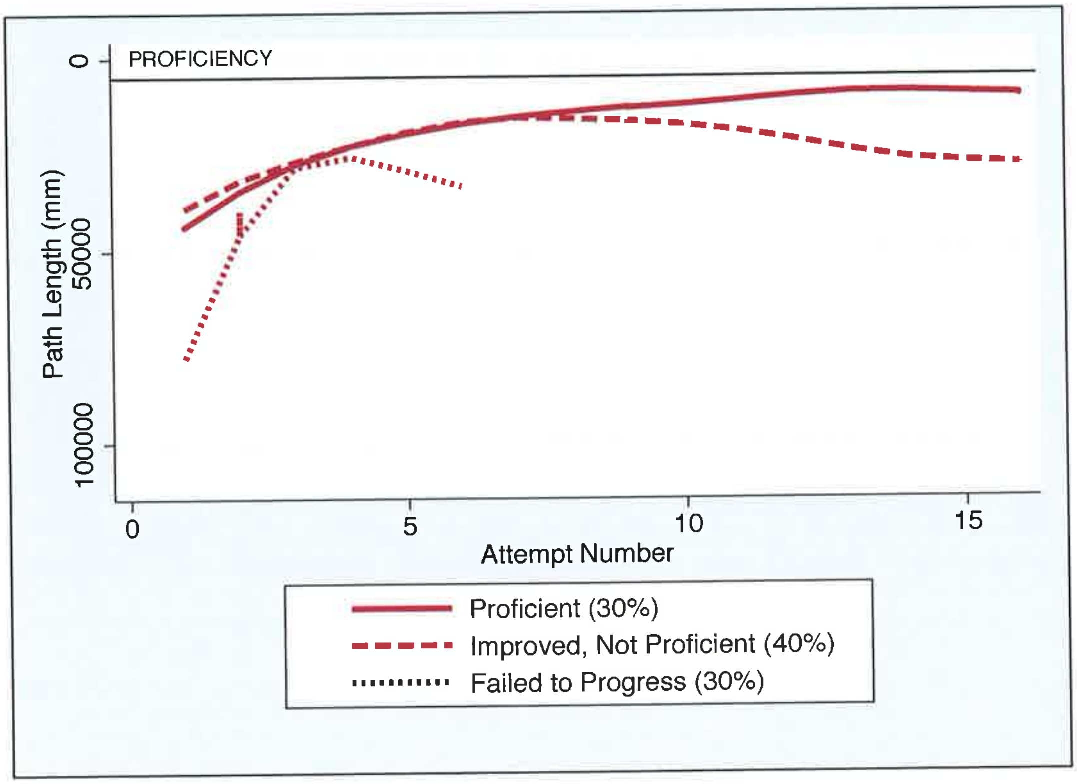


Figure 4.6 Path length scores of medical students with low aptitude in the laparoscopic suturing group

4.4.6 Medical Students Economy of Movement Scores

Smoothness scores for all attempts in both groups are displayed in Figure 4.7. Group A achieved proficiency in a shorter time frame for these two parameters than group B ($p<0.0001$). Again, group B can be further divided into three subgroups for smoothness scores with 30% attaining proficiency, 40% showing improvement but not attaining proficiency and 30% failing to progress, which is shown in Figure 4.8.

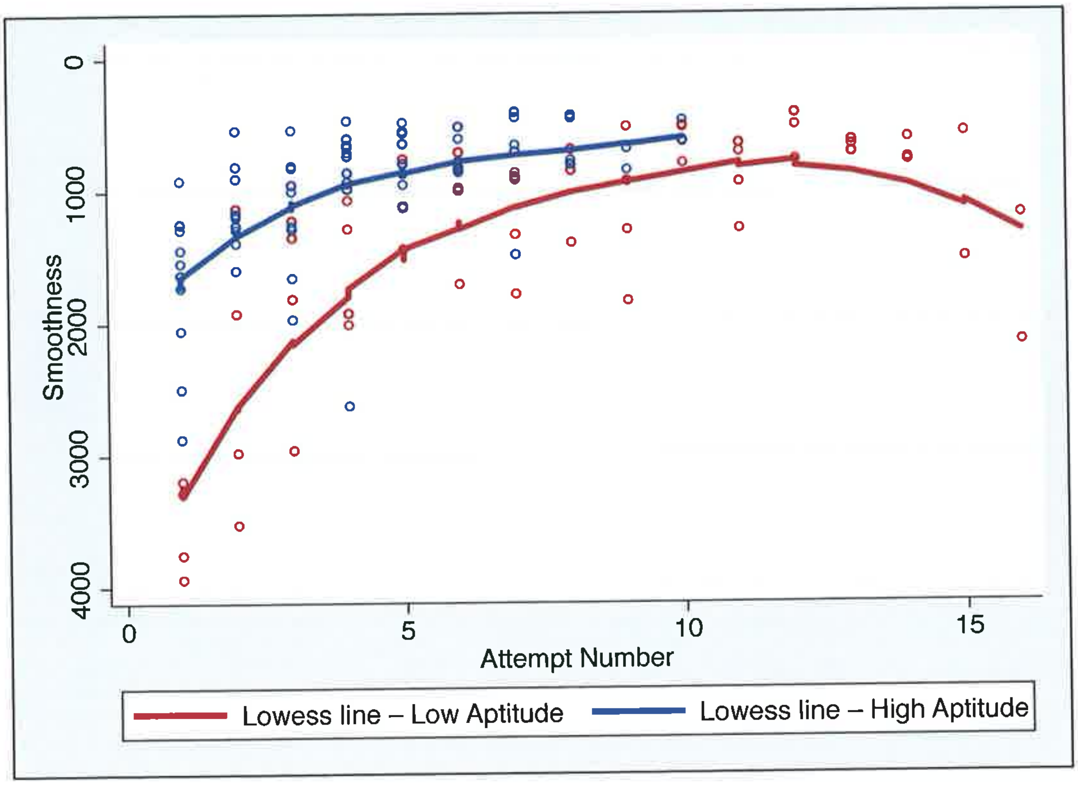


Figure 4.7 Smoothness scores of medical students in the laparoscopic suturing group

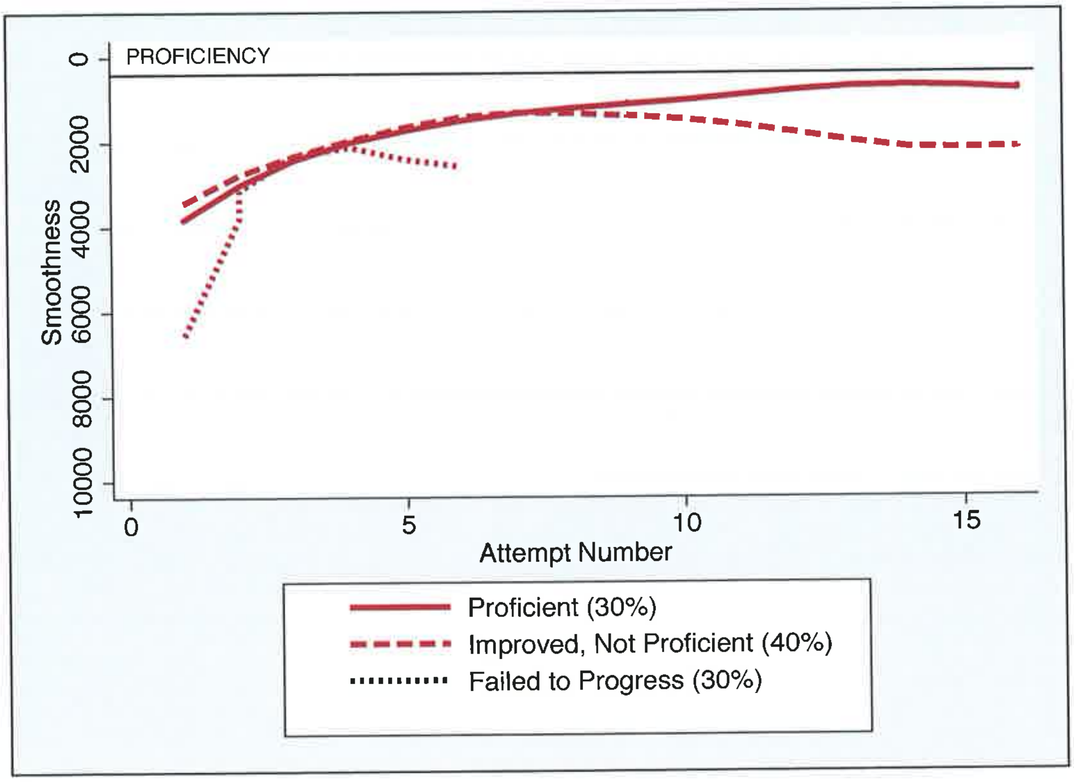


Figure 4.8 Smoothness scores of medical students with low aptitude in the laparoscopic suturing group

4.4.7 Medical Students Time Scores

Time scores for all attempts in both groups are displayed in Figure 4.9. Group A achieved proficiency in a shorter time frame for these two parameters than group B ($p<0.0001$). Again, group B can be further divided into three subgroups for time scores with 30% attaining proficiency, 40% showing improvement but not attaining proficiency and 30% failing to progress, which is shown in Figure 4.10.

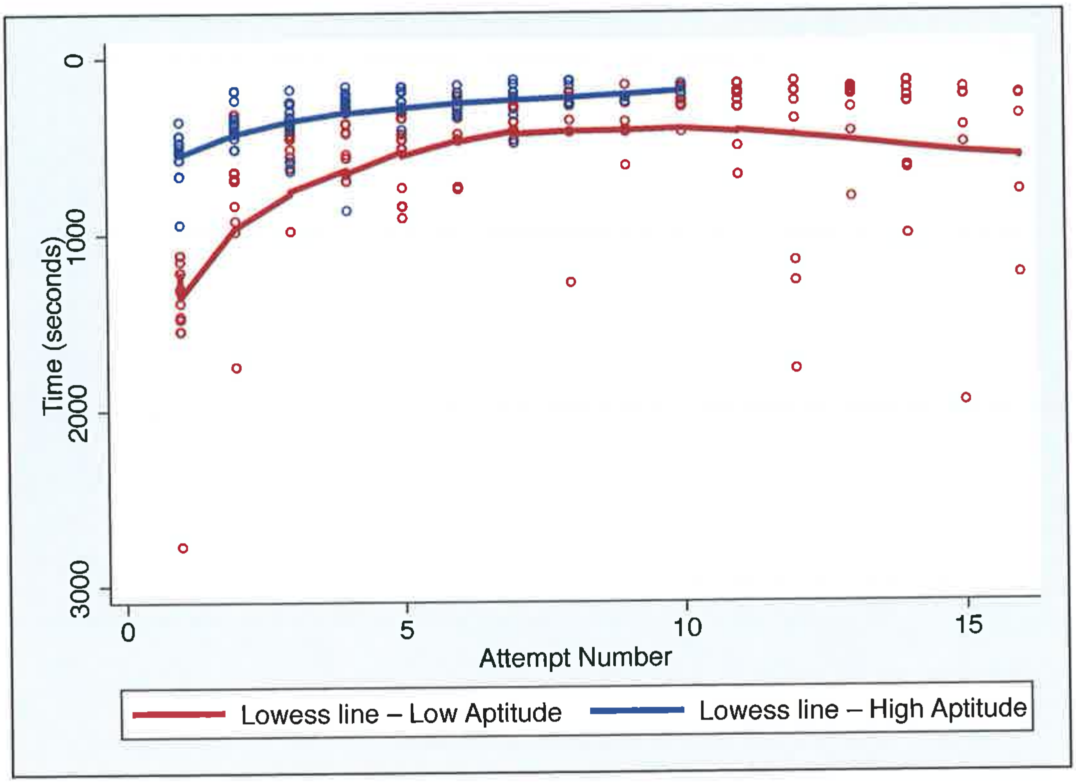


Figure 4.9 Time scores of medical students in the laparoscopic suturing group

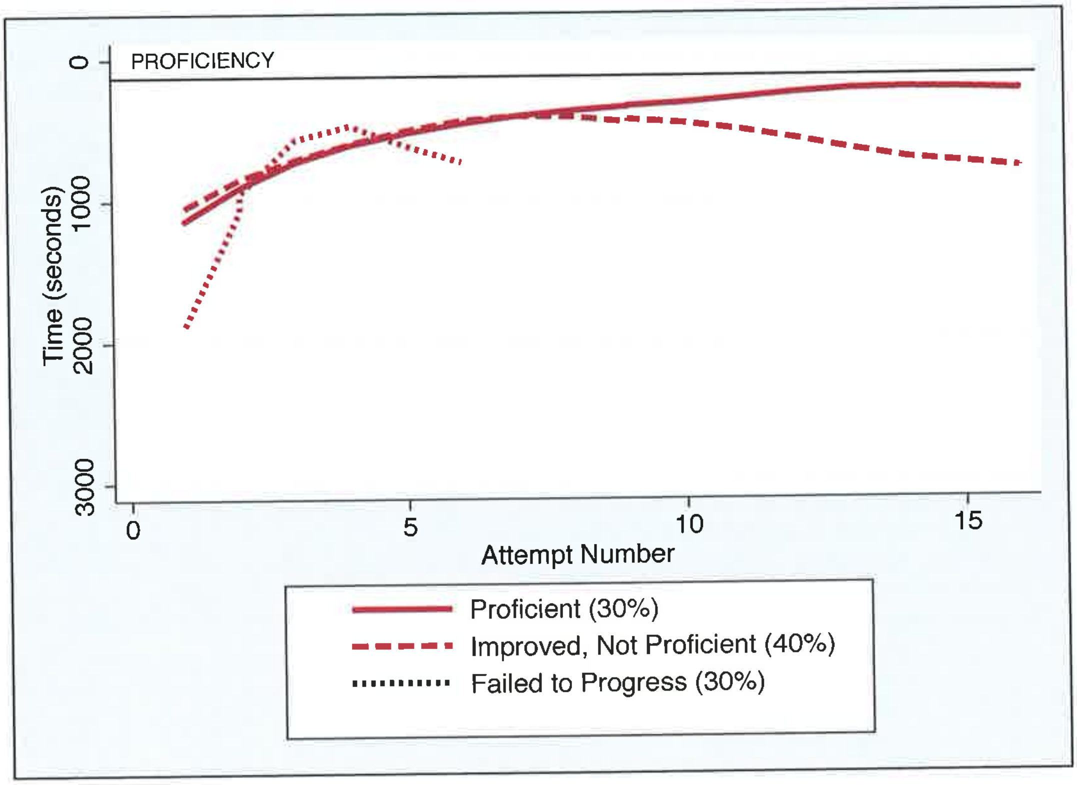


Figure 4.10 Time scores of medical students with low aptitude in the laparoscopic suturing group

4.4.8 Medical Students OSATS and FLS Scores

The baseline subjective scores (OSATS, FLS and error scores) for both groups are shown in Table 4.5; these are the scores from the first attempt for all candidates. Group A achieved significantly better scores compared to group B in all parameters ($p<0.0001$).

OSATS scores for both groups are shown in Figure 4.11 and FLS scores are shown in Figure 4.13 Group A achieved proficiency with lesser attempts ($p<0.001$) than group B. The consistent separation in learning curves among group B is again present with regard to OSATS and FLS scores as depicted in figure 4.12 and 4.14 respectively. Inter-rater reliability between the two assessors was determined using ICC and was found to be 0.96 and 0.94 for FLS and OSATS scores respectively.

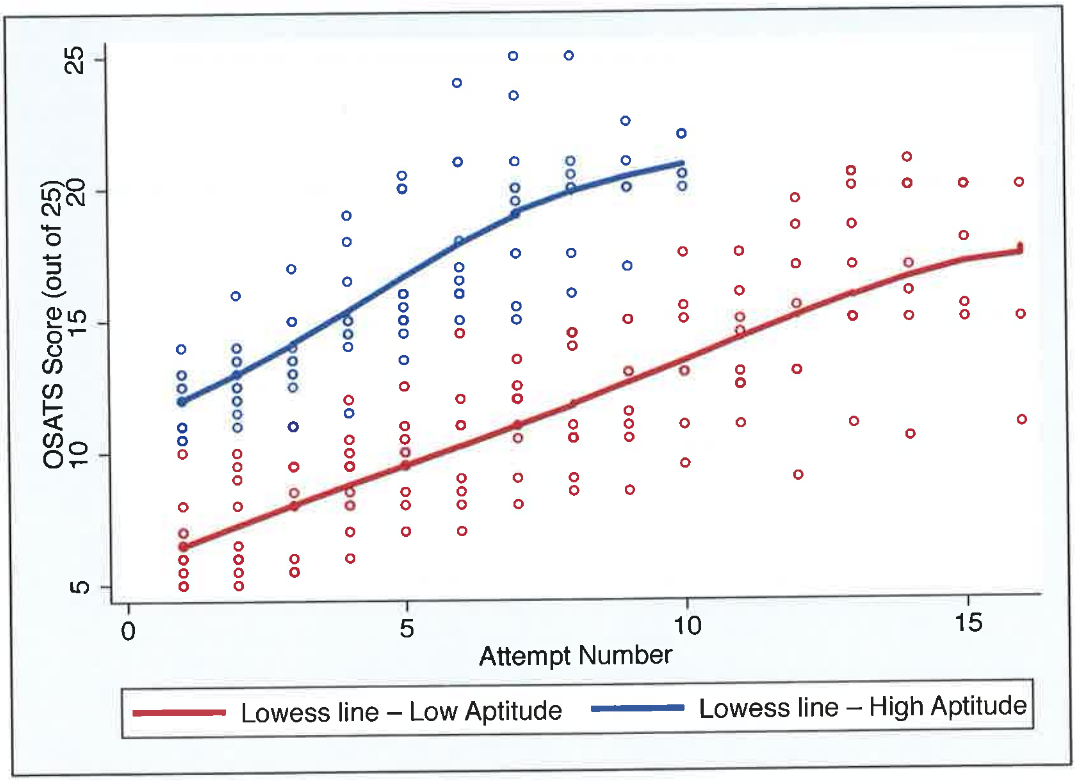


Figure 4.11 OSATS scores of medical students in the laparoscopic suturing group

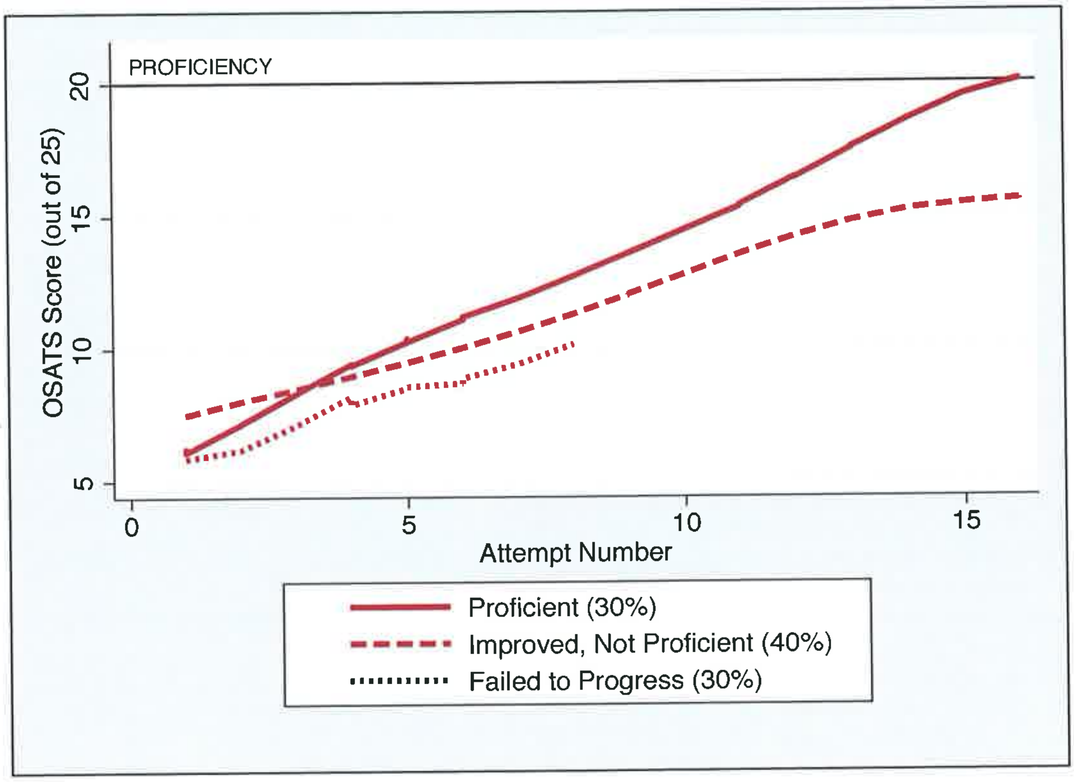


Figure 4.12 OSATS scores of medical students with low aptitude in the laparoscopic suturing group

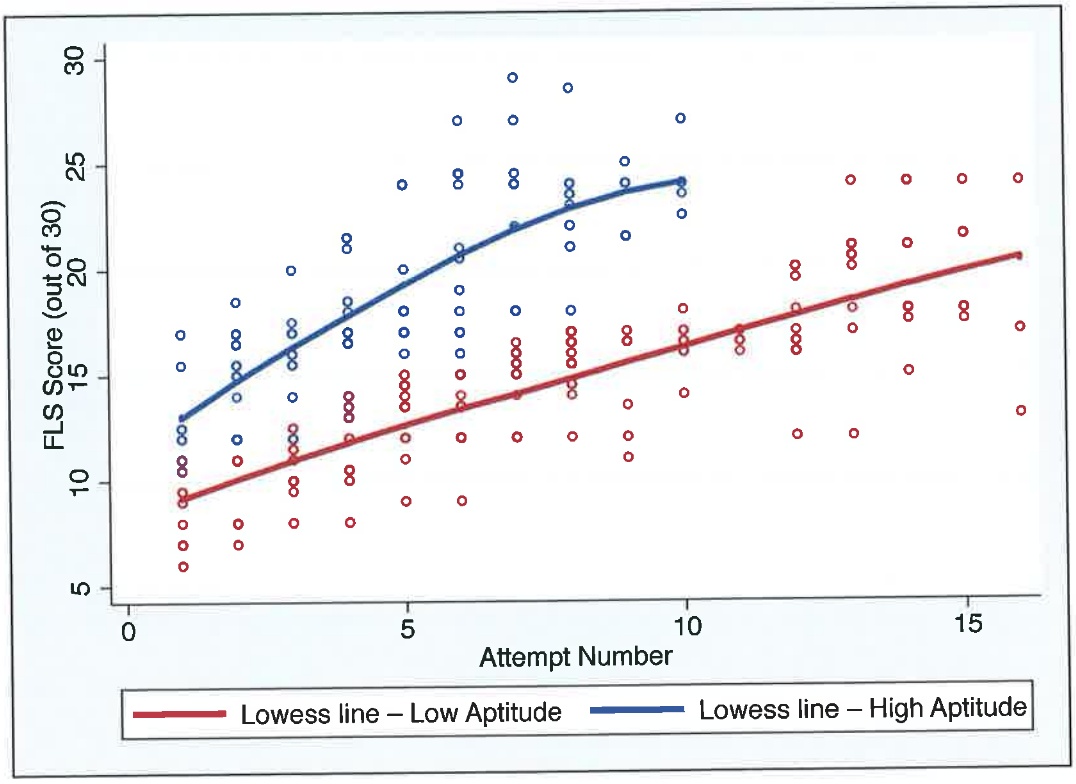


Figure 4.13 FLS scores of medical students in the laparoscopic suturing group

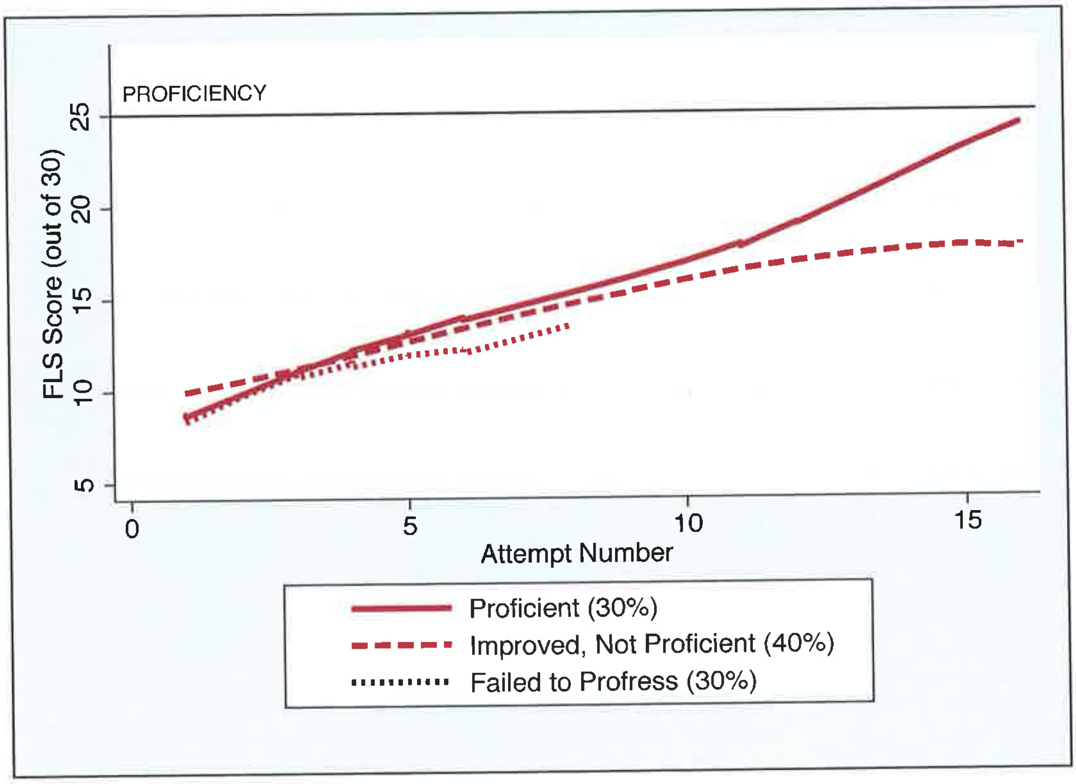


Figure 4.14 FLS scores of medical students with low aptitude in the laparoscopic suturing group

4.4.9 Medical Students Error Scores

The error scores for both groups are shown in figure 4.15. The baseline mean error score was 2.8 in group A compared with 4.7 in group B ($p<0.001$). Inter-rater reliability using was 0.80.

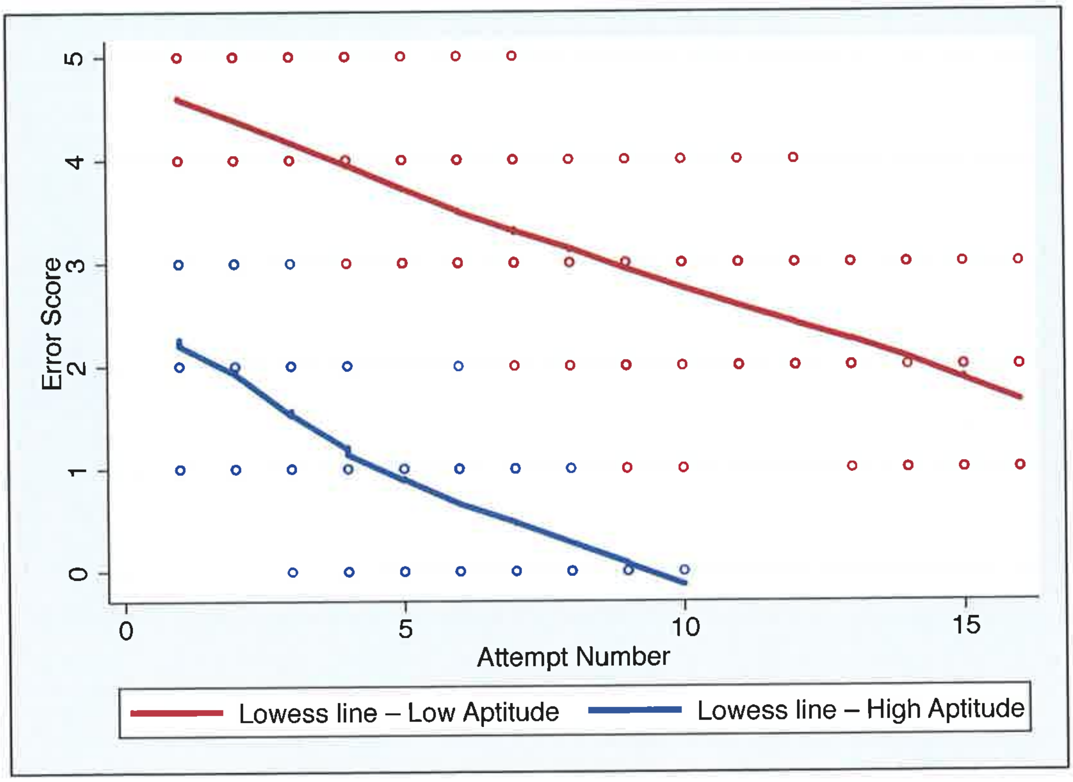


Figure 4.15 Error scores of medical students in the laparoscopic suturing group

4.4.10 Correlation between Aptitude and Baseline Performance in BST's

The results of the correlation between aptitude and baseline performance in surgical trainees is shown in table 4.6. Again, similar to the results seen in Chapter 3 there was a significant correlation between baseline performance and visual spatial aptitude. However the correlation of baseline performance with depth perception aptitude and psychomotor ability was non significant.

Table 4.6 Correlation between aptitude and baseline performance of BST's in laparoscopic suturing

Aptitude (N=10)	Baseline Performance	Correlation (R)	p value
Visual-Spatial	Time	R = 0.690	p = 0.04
	Path Length	R = 0.626	p = 0.07
	Smoothness	R = 0.678	p = 0.04
Depth Perception	Time	R = 0.610	NS
	Path Length	R = 0.462	NS
	Smoothness	R = 0.562	NS
Psychomotor	Time	R = 0.350	NS
	Path Length	R = 0.231	NS
	Smoothness	R = 0.435	NS

4.4.11 Correlation between Aptitude and Attainment of Proficiency in BST's

We assessed the relationship between aptitude and ability to reach the predefined proficiency scores in laparoscopic suturing. The results demonstrated that depth perception aptitude tested impacted positively on the ability of subjects to achieve proficiency (Table 4.7)

Table 4.7 Correlation between aptitude and attainment of proficiency of BSTs in laparoscopic suturing

Aptitude (N = 10)	Proficiency	Correlation (R)	p value
Visual-Spatial	Number of Attempts	R = 0.614	p = 0.07
Depth Perception	Number of Attempts	R = 0.757	p = 0.01
Psychomotor	Number of Attempts	R = 0.597	p = 0.08

4.4.12 BST's Metric Scores

The surgical trainees path length, economy of movement and time scores are shown in figures 4.16, 4.17 and 4.18 respectively. They achieved proficiency after an average of 7 attempts (range 6-9). As typically seen with learning curves the increase in retention of information is sharpest after the initial attempts.

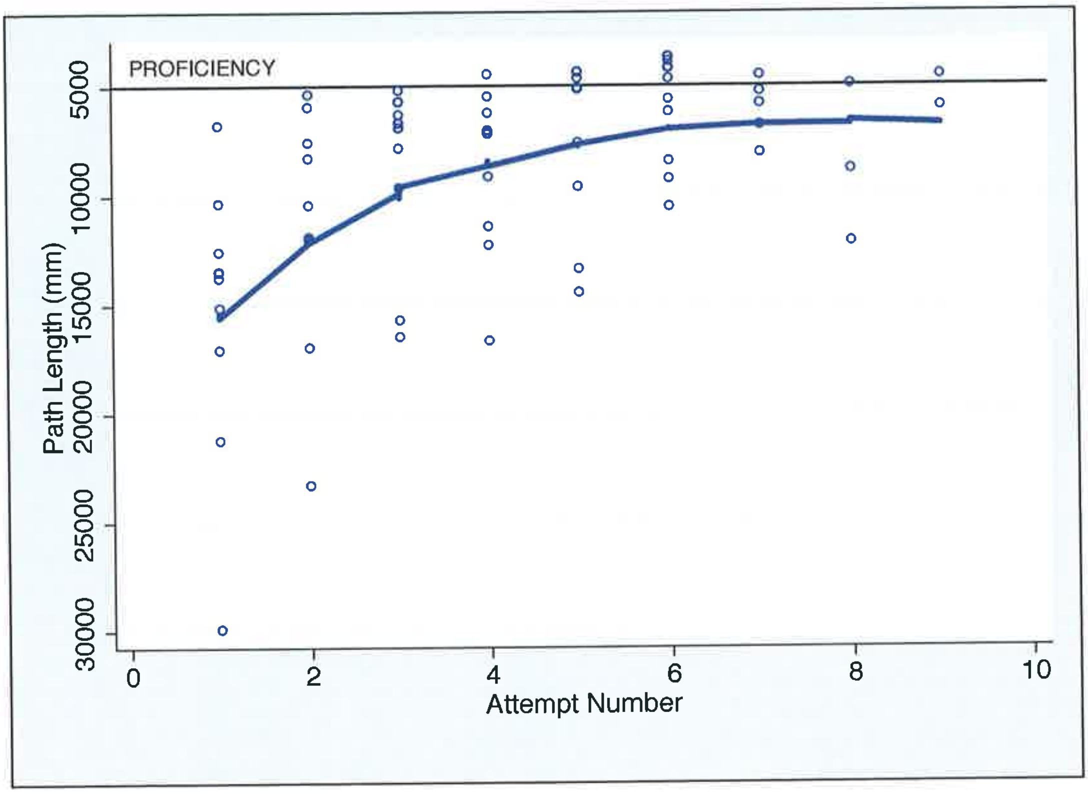


Figure 4.16 BST path length scores in the laparoscopic suturing group

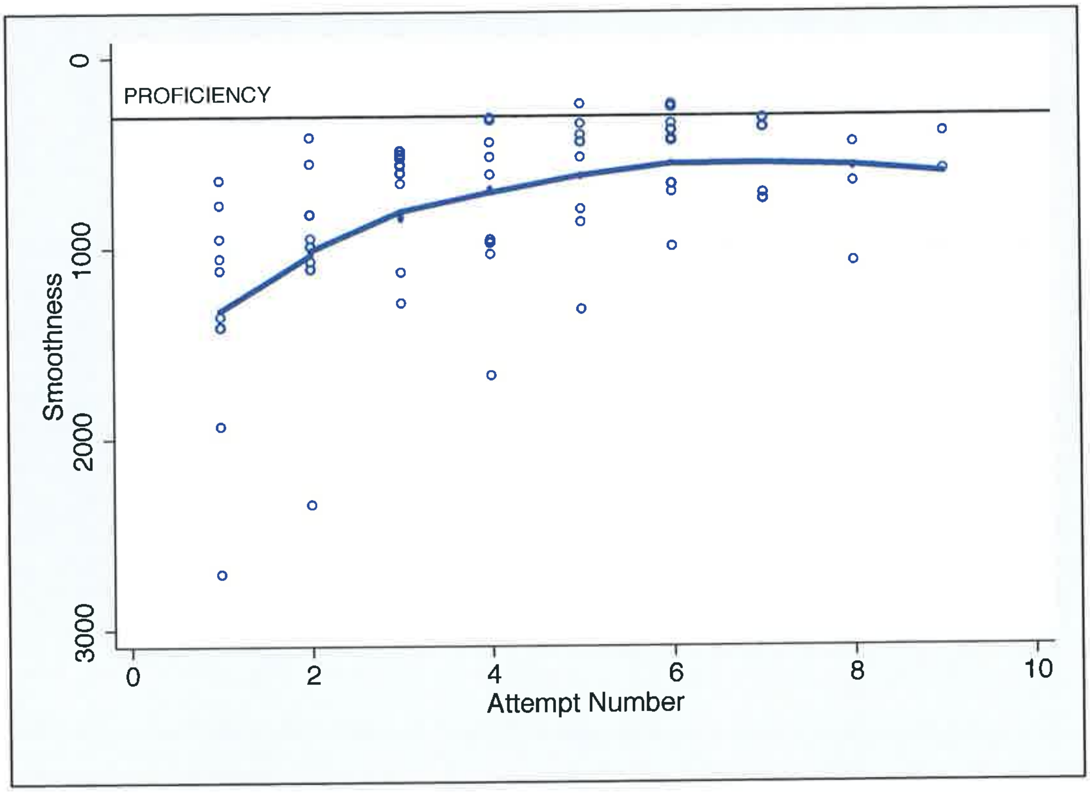


Figure 4.17 BST smoothness scores in the laparoscopic suturing group

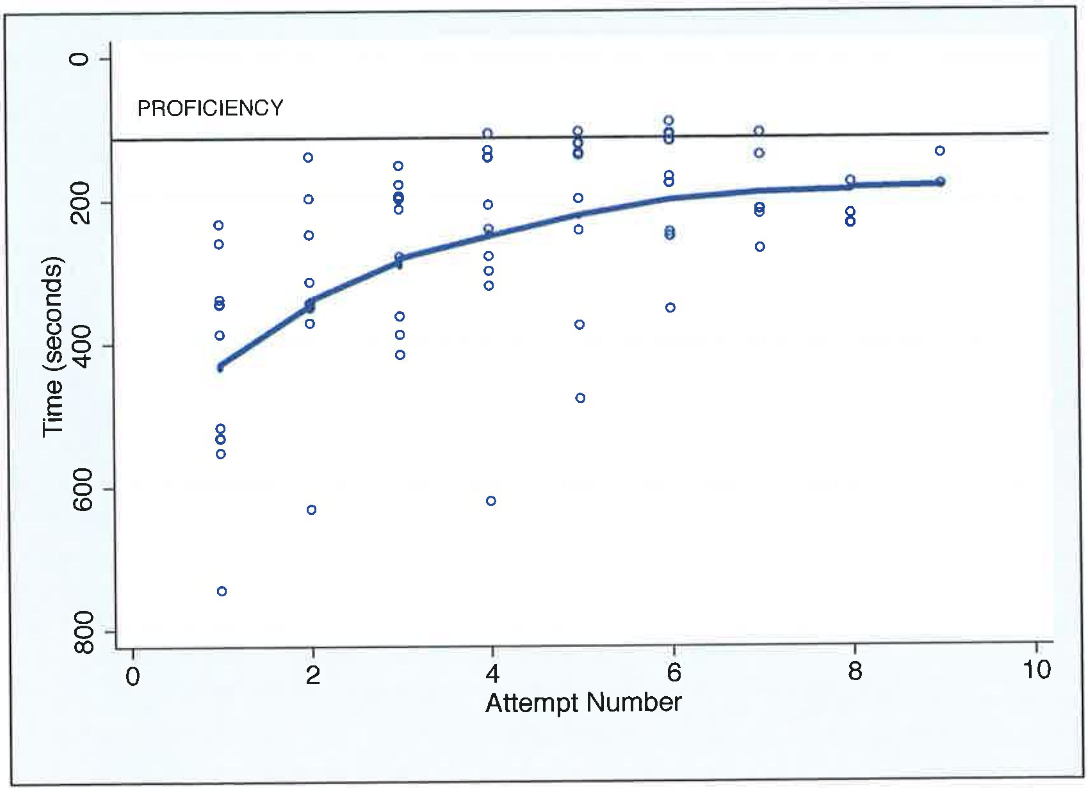


Figure 4.18 BST time scores in the laparoscopic suturing group

4.4.13 BST's Subjective Scores

OSATS scores are shown in Figure 4.19, the average baseline score was 13 compared to 22 at achievement of proficiency. FLS scores are illustrated in Figure 4.20. The mean baseline score was 12 compared to 24 at achievement of proficiency. Inter-rater reliability between the two assessors who rated the laparoscopic suturing performances was determined using Cronbach's alpha and was found to be 0.95 and 0.98 for OSATS and FLS scores respectively.

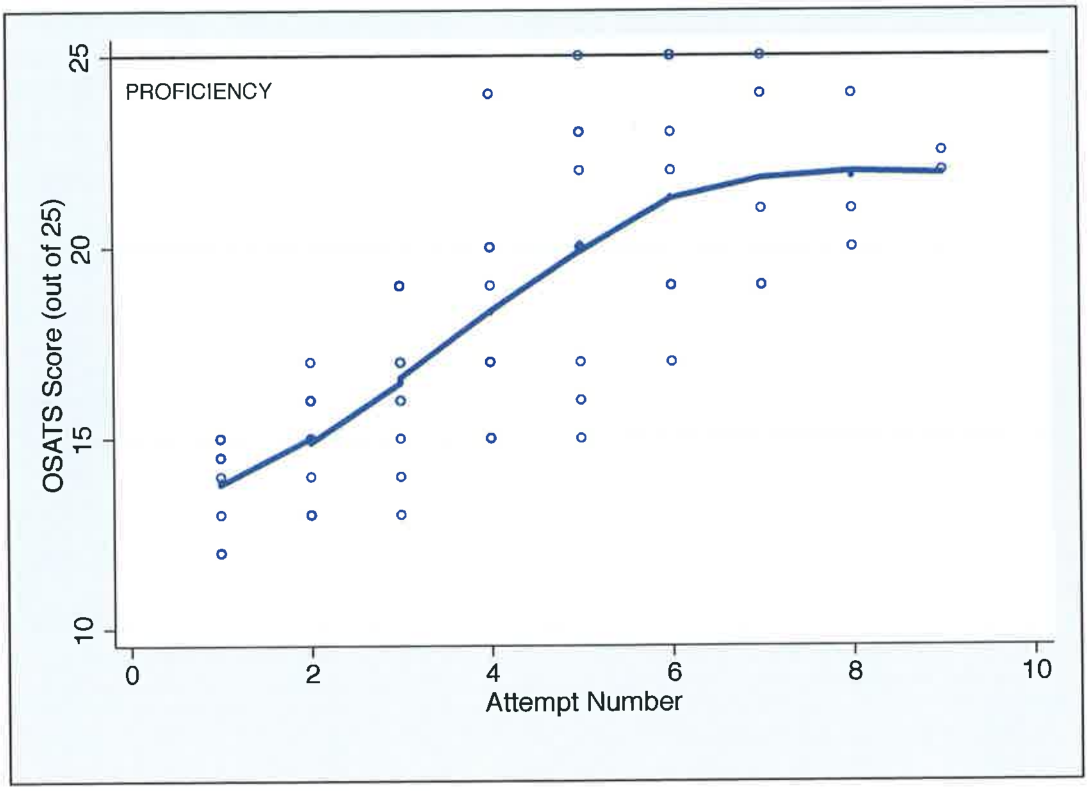


Figure 4.19 BST OSATS scores in the laparoscopic suturing group

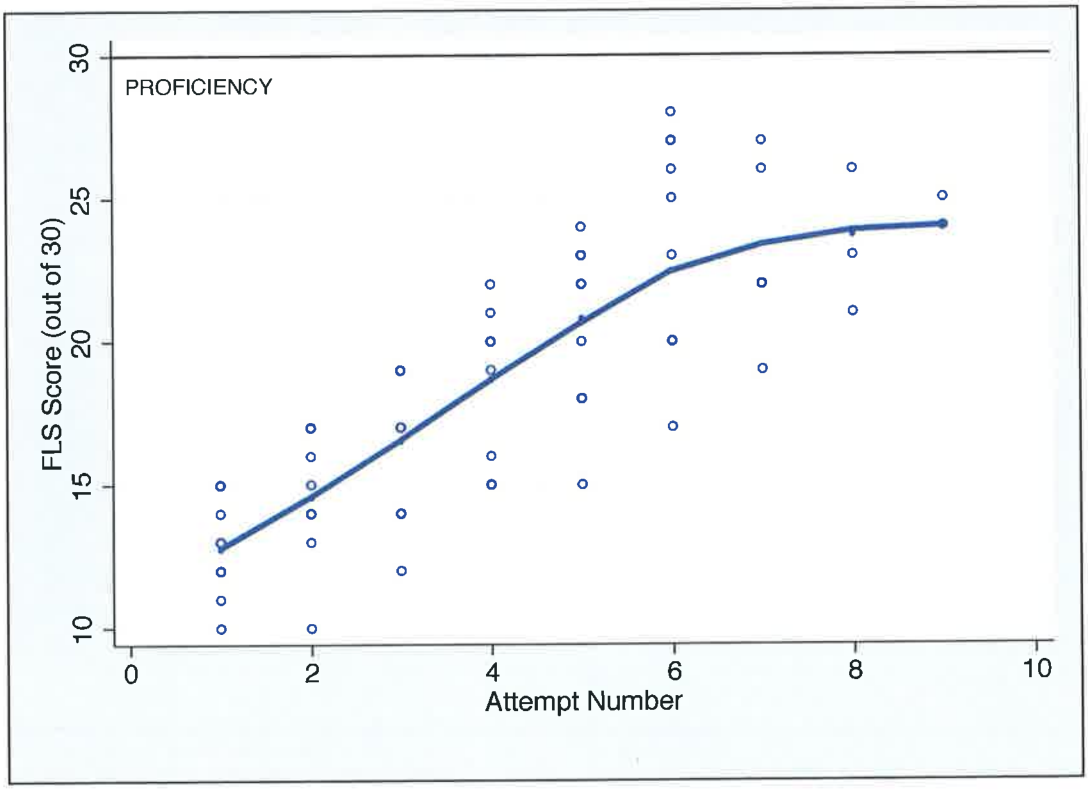


Figure 4.20 BST FLS scores in the laparoscopic suturing group

The error scores are shown in Figure 4.21. The baseline mean error score was 2.3. Inter-rater reliability was 0.86.

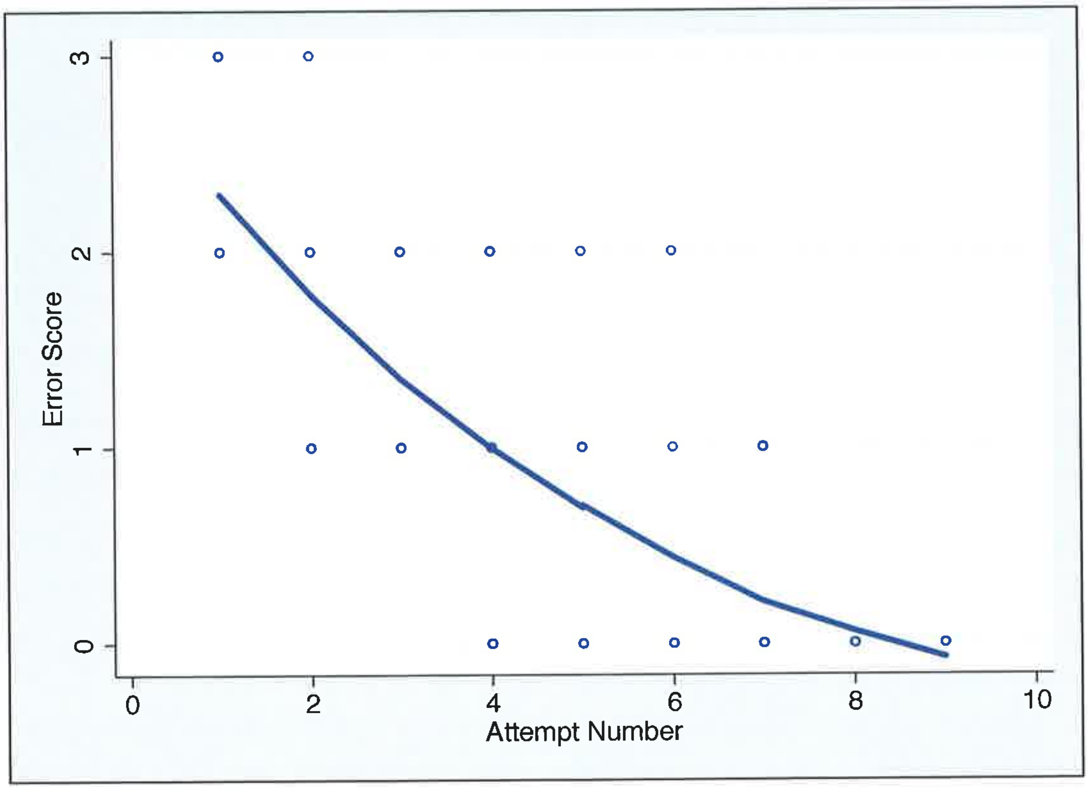


Figure 4.21 BST error scores in the laparoscopic suturing group

4.4.14 Prediction of the Learning Curve for Laparoscopic Suturing

Significant differences were found between the average expert score and the ten surgical trainees score for the first simulated laparoscopic task performed. Scores improved with each subsequent procedure and reached expert scores on an average of 7 attempts. Significant differences were found for time, path length and smoothness (p=0.04), OSATS and FLS scores (p=0.004) and error score (p=0.003) when attempt 1 results were compared with attempt 7. These results are shown in table 4.8

Table 4.8 Baseline and final attempt assessment scores of BST's in laparoscopic suturing

	Attempt 1	Attempt 7	p value (Wilcoxon signed-rank test)
Time, s	422	188	0.04
Path Length, mm	15580	6076	0.04
Smoothness	1327	505	0.04
OSATS (out of 25)	13	22	0.004
FLS (out of 30)	12	24	0.004
Error Score (out of 5)	2.3	0	0.003

4.5 Discussion

It is clear that distinct learning curves for laparoscopic suturing can be mapped based on fundamental ability. This experiment has shown that there is a wide disparity in attaining proficiency in groups with differing aptitude.

Surgical training bodies have many objectives, and one of their most important aims is to ensure that candidates excel in their chosen field. Several specialties' demand advanced technical skills specifically in the minimally invasive environment. Based on our study findings, candidates who possess good fundamental ability are more likely to succeed in their training pathway within a given timeframe.

The current literature shows that residents with higher visual-spatial scores perform significantly better than those with lower scores (Wanzel et al., 2002, Wanzel et al., 2003). Our study supports these findings as we have shown that candidates with high aptitude scores have a superior baseline in all parameters and attain proficiency quicker.

Grantcharov et al (Grantcharov and Funch-Jensen, 2009) previously identified four types of learning curves, which varied based on psychomotor ability. Our data strongly supports these findings. We also found that there is the proportion of students who fail to progress along the learning curve.

In the previous chapter we demonstrated that 25% of candidates with low aptitude were unable to achieve proficiency in laparoscopic appendectomy despite repeated attempts. This study shows that the proportion is higher, when a complex task such as laparoscopic suturing is being attempted. Only 30% of the candidates with low aptitude were able to achieve proficiency in the laparoscopic suturing task, implying that aptitude plays an important role in learning advanced laparoscopic skills.

For the BST group we demonstrated the learning curve required by trainee surgeons to reach technical proficiency in laparoscopic suturing using simulation-based training. An average of 7 attempts was required to reach benchmark goals and all objective metrics ($p=0.04$) and subjective scores ($p=0.004$) improved from baseline to proficiency.

In the BST group we found that visual spatial ability significantly correlated with superior baseline performance in completing this laparoscopic task. We also found that depth perceptual ability correlated significantly with the number of attempts required to attain proficiency in laparoscopic suturing.

In conclusion, high aptitude predicts a faster learning curve and improved performance in laparoscopic suturing. A significant number of candidates with low innate ability are unable to reach proficiency despite repeated practice. This study supports the concept of using objective selection processes based on aptitude to select suitable trainees who are likely to flourish in the field of surgery if selected.

Chapter 5

The Role of Self-Selection into Surgery Based on Fundamental Ability

5.1 Introduction

The cost and time commitment of training a surgeon is considerable. Currently in Ireland, surgical trainees must undergo 2 years of basic surgical training (BST) and one year of basic specialty training (BSpT) prior to becoming eligible to apply for a further 6 years of higher surgical training (HST) (www.rcsi.ie). A similar program is run in the UK (www.iscp.ac.uk). In North America, candidates can either go straight into a general surgery residency or else complete an independent model residency which is a general surgical residency followed by a specialist surgical specialty training program (www.facs.org). No matter where training is delivered it is expensive and lengthy and the trainees should ideally be those best suited to the specialty.

The RCSI has developed a fair and transparent selection process for HST selection into general and plastic surgery, which has been implemented since 2006 (Carroll et al., 2009). Further to this, all applicants for HST selection into surgery undergo assessment for fundamental aptitudes and abilities. A prospective database has been maintained since 2007 of all trainee scores and is being correlated to current clinical performance (Gallagher et al., 2009). Correlating current clinical performance against past aptitude and ability scores will hopefully aid and assess the selection process of HST and identify applicants who might struggle during training and thereafter.

It was noted in the previous two chapters that none of the BST's had similar scores to the low innate ability group in either experiment. Therefore it could be possible that the trainees are self-selecting themselves into a career in surgery based on innate ability.

Prior work has been done in our institution into the role of aptitude in the acquisition of basic and advanced minimally invasive skill sets (Nugent et al., 2012a) and microsurgical skills acquisition (Nugent et al., 2012c); therefore we feel aptitude to be an important indicator of surgical performance. The self-selection process of candidates with innate psychometric ability into a given field of surgery is likely and must be considered.

5.2 Objectives

5.2.1 Hypothesis Underlying the Objectives

This study was undertaken in order to compare the aptitude scores of all applicants to the higher surgical training scheme in general and plastic surgery and a population of surgical naïve medical students who have an interest in pursuing a career in surgery. Our hypothesis was that the majority of applicants into higher surgical training would possess good fundamental abilities necessary to become competent surgeons and will thereby self-select themselves appropriately.

5.2.2 Detailed Objectives

Objective 1. To assess the difference in aptitude between general surgery HST applicants and medical students. We aimed to compare the all aptitude scores (visual spatial, depth perception and psychomotor) of general surgery HST applicants to a cohort of medical students

Objective 2. To assess the difference in aptitude between plastic surgery HST applicants and medical students. We aimed to compare the all aptitude scores (visual spatial, depth perception and psychomotor) of plastic surgery HST applicants to a cohort of medical students

Objective 3. To investigate if there was a proportion of HST applicants who had a suboptimal aptitude score.

It was hypothesisied that not all surgical candidates self selected into surgery appropriately, therefore the proportion of HST applicants who scored below the mean aptitude scores of the surgical novice cohort was calculated for each aptitude tested.

Objective 4. To investigate if there was a difference in mean aptitude scores between the applicants who were sucessful and unsucessful at gaining entry to the higher surgical training scheme.

In order to determine is there was any aptitude difference between candidates who were successful and unsuccessful into HST, an analysis of the scores between the two groups was performed.

5.3 Materials and Methods

5.3.1 Participant Recruitment

Candidates eligible for Higher Surgical Training in Ireland can apply to the Royal College of Surgeons through a standardised process (Carroll et al, 2009; Gallagher et al, 2008). Candidates with the highest scores are then shortlisted for interview and further assessment.

All candidates shortlisted for HST selection in general and plastics and surgery have been undergoing assessment for fundamental aptitude since 2007. They undergo a series of validated tests in order to assess for visual spatial and perceptual ability. The grooved pegboard test, which assesses psychomotor ability, was introduced in 2010. This data is collected and stored in a confidential and secure database.

The Research Ethics Committee of RCSI granted ethical approval. All shortlisted candidates gave informed and written consent to have any data collected as part of the assessment process analysed and used for research purposes. A prospective database was set up to gather all data collected from the short-listed assessments between the years 2007-2013.

A group of medical students (year 1-6) were recruited to take part in this study via class emails that were circulated. They volunteered to take part in this study as a comparison group for the HST applicants and underwent the same aptitude testing in the same manner. These students all had an interest in pursuing a career in surgery. They were enrolled into the study on a first come first served basis and gave written informed consent to take part in the study. It was made clear to them that all data collected was to be anonymous and was to be stored in a secure and anonymous fashion.

5.3.2 Participant Demographics

A total of 300 candidates underwent aptitude testing for the purpose of this study, 146 medical students and 154 HST applicants. 113 of the HST candidates applied to general surgery and 41 to plastic surgery. Demographic details are shown in Table 5.1.

Table 5.1. Demographics of HST applicants

	HST Applicants (Plastic Surgery)			HST Applicants (General Surgery)			Medical Students		
	Male	*Female*	*p*	*Male*	*Female*	*p*	*Male*	*Female*	*p*
Gender	17	24	-	70	43	-	67	79	-
Age Range	29-36	29-35	-	29-36	29-36	-	18-32	18-30	-
Visual Spatial									
Mean	66.8	54.5	0.08	62.9	63	0.88	51.7	44.6	0.2
St. Dev.	± 20	± 17		± 21	± 17		± 19	± 23	
Depth Perception									
Mean	93.4	87.3	0.06	89.8	84.7	0.02	79.8	73.8	0.3
St. Dev.	± 4.5	± 9.6		± 14	± 18		± 12	± 15	
Psychomotor									
Mean	58.5	57.4	0.52	62	56.6	0.05	68.6	65.2	0.4
St. Dev.	± 5.6	± 6.6		± 10	± 7		± 12	± 15	

5.3.3 Aptitude Assessment

Visual-spatial, depth perception, and psychomotor aptitudes have been previously demonstrated to be associated with surgical technical skill using the various tests we chose for this study (Buckley et al., 2014a, Enochsson et al., 2006, Hassan et al., 2007, Van Herzeele et al., 2010). All candidates were assessed for visual-spatial, depth perception and psychomotor aptitude as previously described in Chapter 2. All candidates in the HST groups and student group underwent aptittude testing in the manner and under the same conditions.

Two different aspects of visual spatial aptitude were examined, spatial orientation and spatial scanning. Spatial orientation was assessed using the card rotations test and spatial scanning was assessed using the map planning test. Both of these tests were taken from the Kit of Factor Referenced Cognitive Tests (Ekstrom et al, 1976). Perceptual ability was assessed using PicSOr (Pictorial Surface Orientation) as developed by Cowie in 1998 (Cowie, 1998). Psychomotor aptitude was assessed using the Grooved Pegboard (Dikmen et al, 1999).

5.3.4 Comparison of Scores

Scores for visual spatial, depth perception and psychomotor aptitude were calculated for both the group of medical students and the applicants to the higher general and plastics surgical scheme. The mean scores of both the HST applicant groups were compared to the student population group using the k-sample equality of medians test. Further to this, the percentage of HST who performed below the mean performance score of the medical student was calculated. An analysis of the scores between the candidates who were successful and unsuccessful into HST was also performed.

5.3.5 Statistical Analysis

Statistical analysis was performed using Stata 12.0. The mean aptitude scores of the HST applicants and medical students were compared using the k-sample equality medians test, $p<0.05$ was considered statistically significant. Non-parametric testing was performed as none of the data was shown to have a normal distribution using the Shapiro wilk test (visual spatial scores $p<0.0001$; depth perception $p<0.0001$; psychomotor 0.004).

5.4 Results

5.4.1 Participant Demographics

A total of 300 candidates underwent aptitude testing for the purpose of this study, 146 medical students and 154 HST applicants. 113 of the HST candidates applied to general surgery and 41 to plastic surgery. Demographic details are shown in Table 1.

5.4.2 Comparison of Visual-Spatial Scores

The mean score for the general surgical HST applicants was 62.3% and the mean score for the plastic surgery HST applicants was 59.7 compared with 31.1% for the medical students (Figure 5.1 and 5.2 respectively).

The general HST applicants achieved a significantly higher score in the card rotation, cube comparison and map-planning test than the medical students (Figure 5.3). The applicants to the plastic surgery higher surgical training scheme also achieved higher scores in all aspects of visual spatial aptitude but only the cube comparison and map planning test were statistically significant (Figure 5.4).

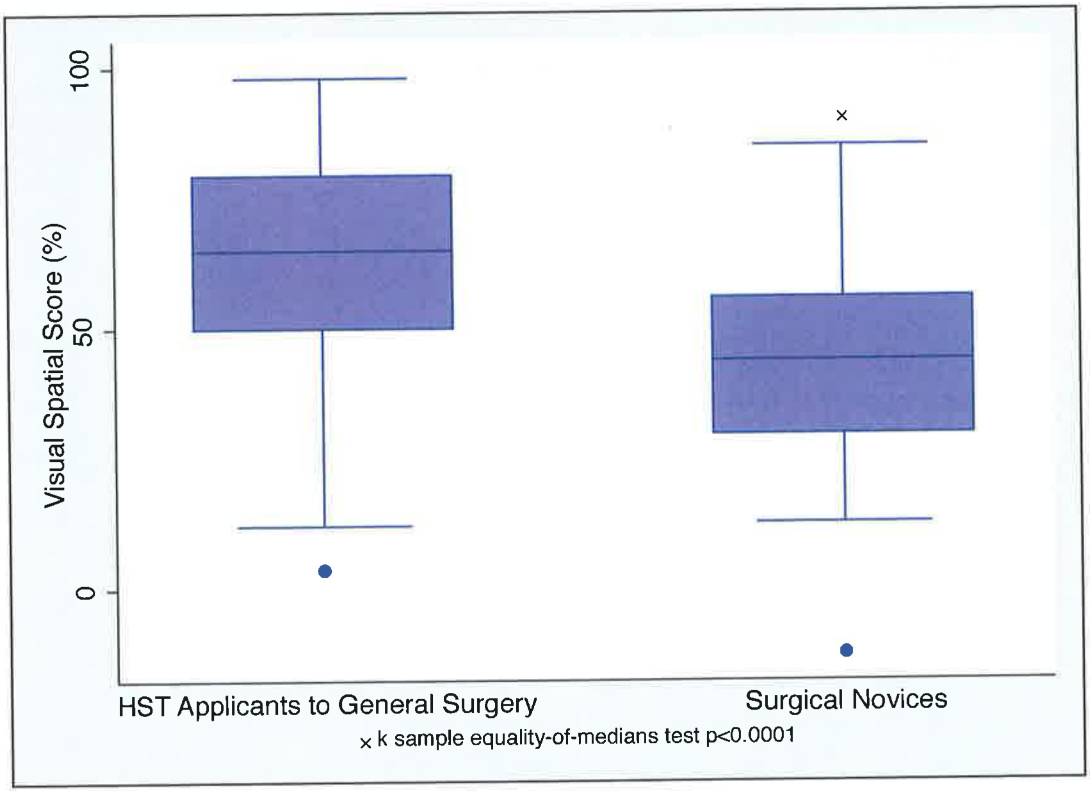


Figure 5.1 Visual spatial scores of general surgery HST applicants and medical students

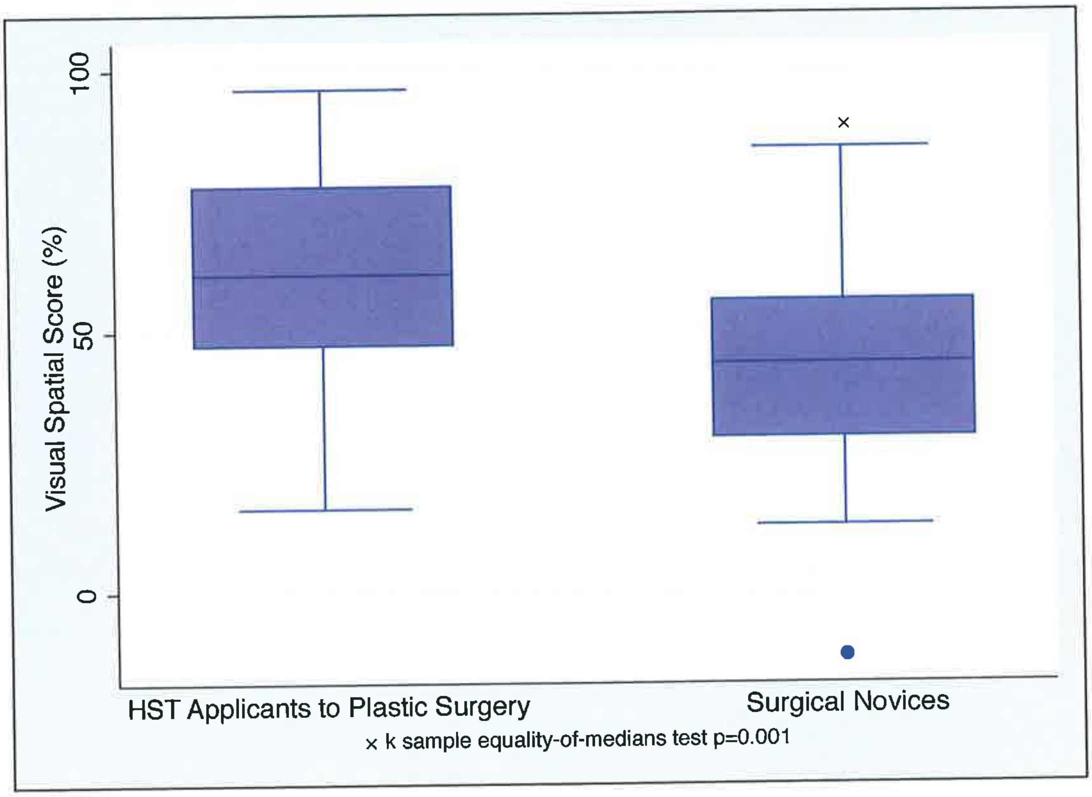


Figure 5.2 Visual spatial scores of plastic surgery HST applicants and medical students

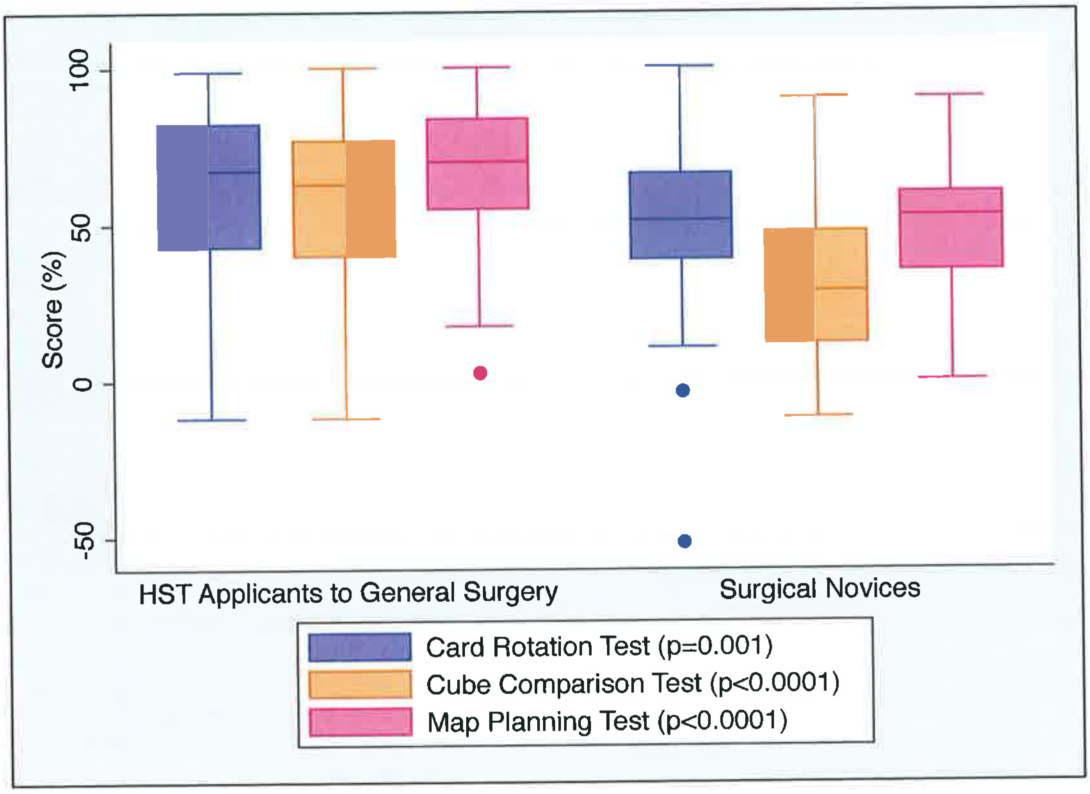


Figure 5.3 Breakdown of visual spatial aptitude scores between general surgery HST applicants and medical students

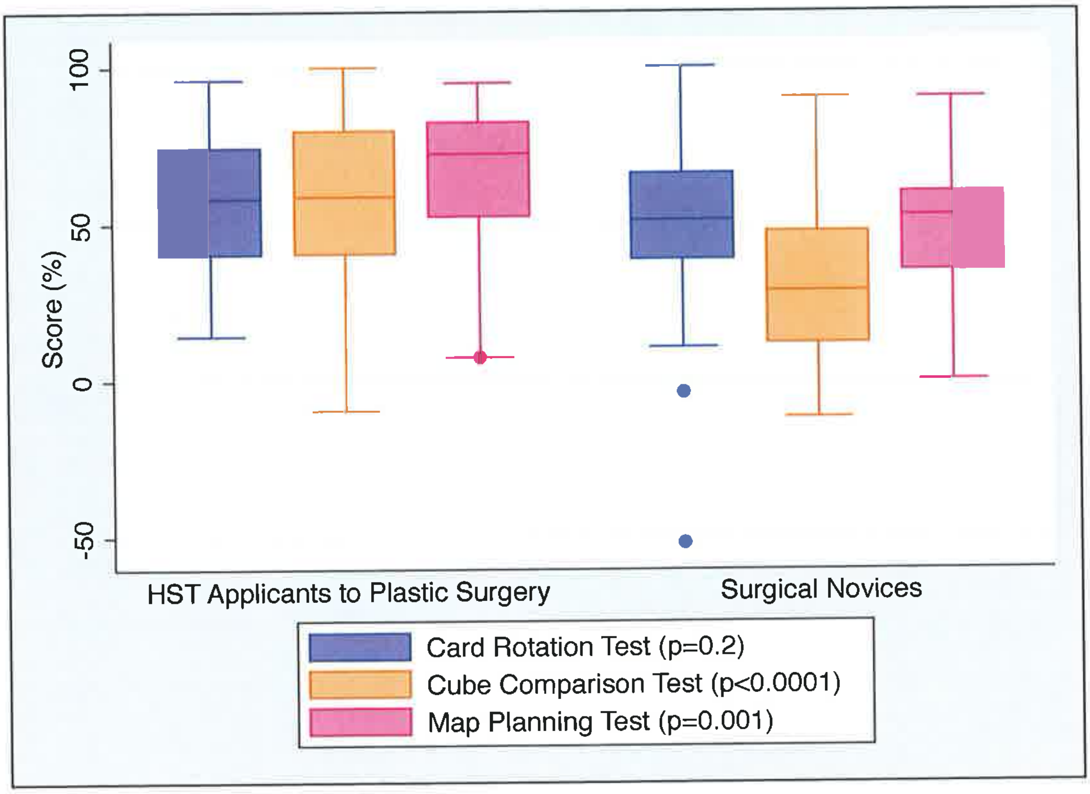


Figure 5.4 Breakdown of visual spatial aptitude scores between plastic surgery HST applicants and medical students

5.4.3 Comparison of Depth Perception Scores

The mean correlation picsor score for general HST applicants was 87.7 compared to 76.6 in the medical student group ($p<0.0001$) (Figure 5.5). The HST applicants to plastics surgery also had higher mean correlation scores; 89.7 compared to 76.6 respectively ($p<0.0001$) (Figure 5.6).

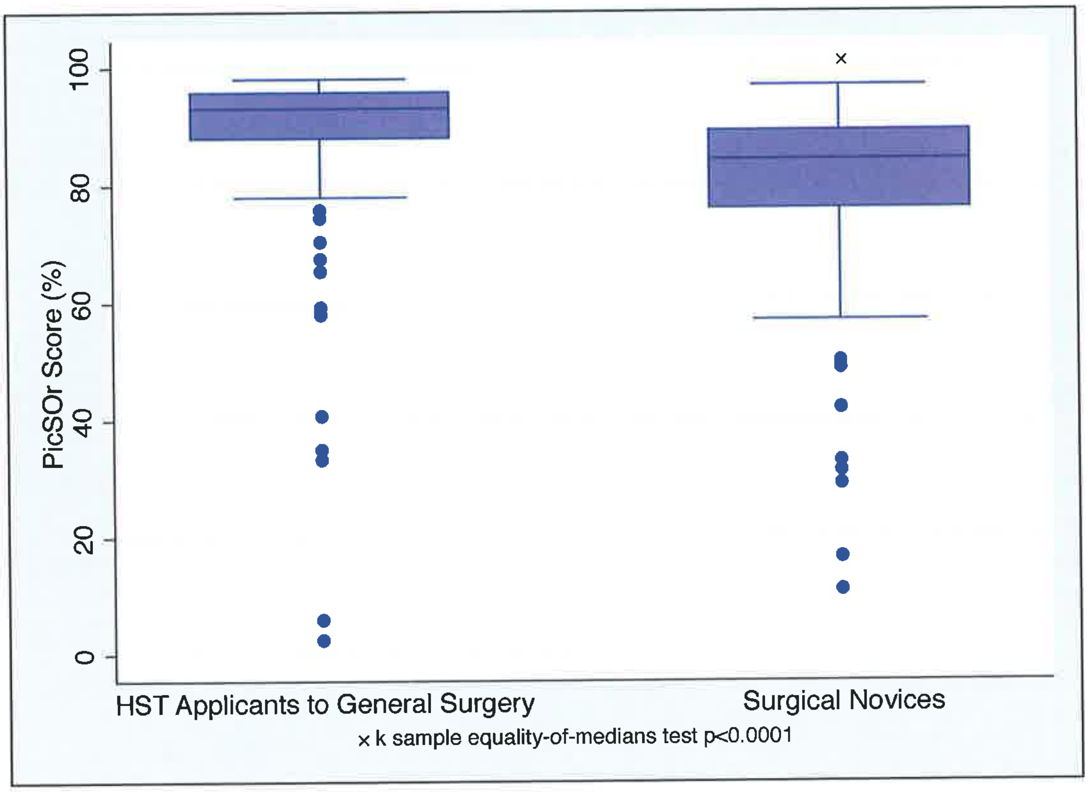


Figure 5.5. Perceptual scores of general surgery HST applicants and medical students

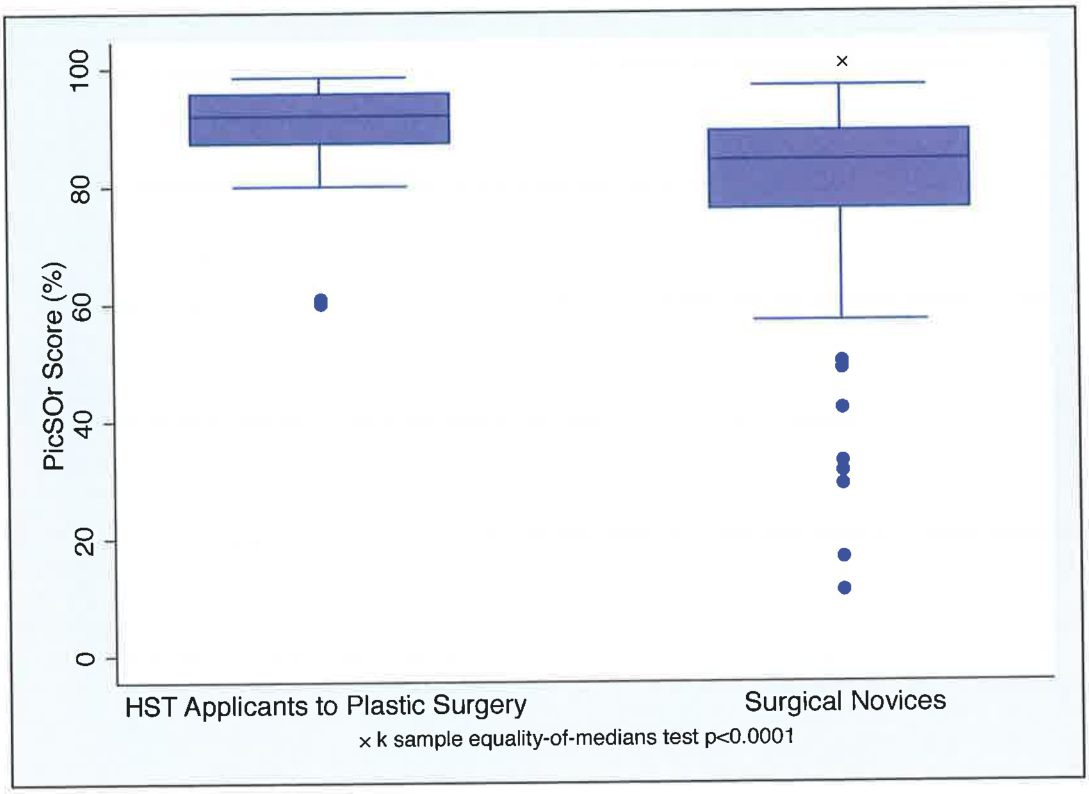


Figure 5.6. Perceptual scores of plastic surgery HST applicants and medical students

5.4.4 Comparison of Psychomotor Scores

Both HST groups performed the pegboard test faster than the medical students. The mean average score was 59.2 seconds in the general HST group and 57.9 in the plastic HST group. The medical students had a mean score of 64.2 seconds. Figures 5.7 and 5.8 display the general and plastic surgery HST group's scores in comparison to the novice group respectively.

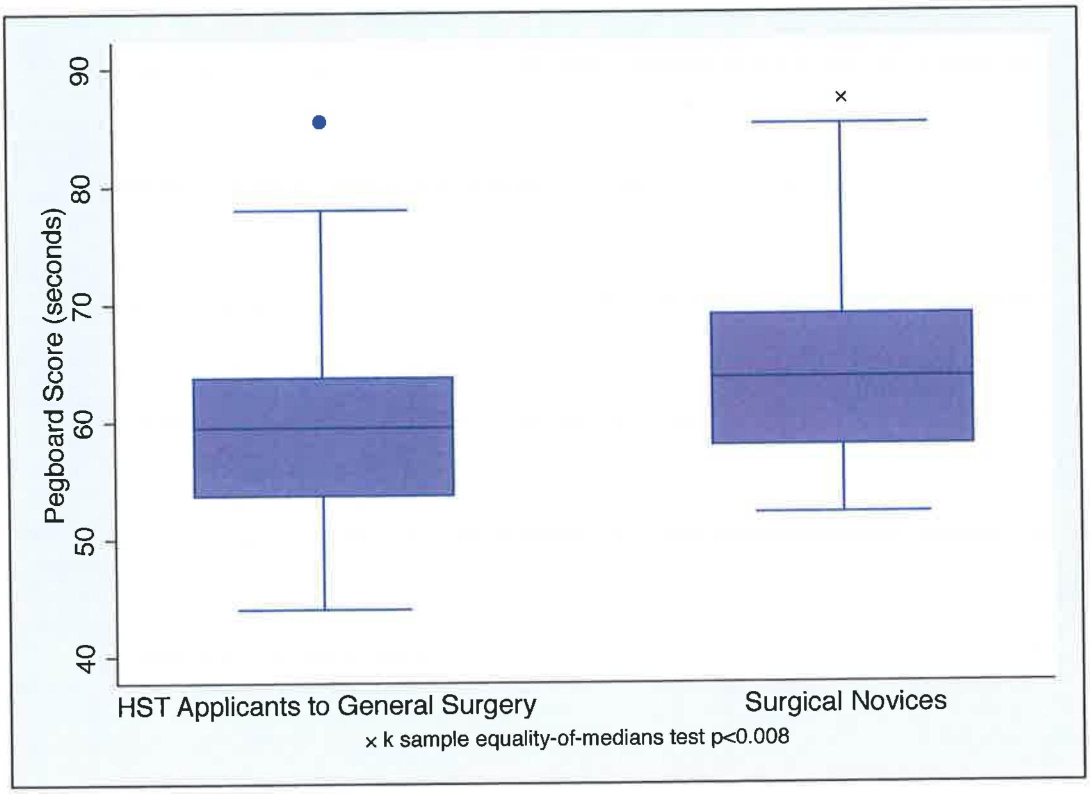


Figure 5.7 Psychomotor scores of general surgery HST applicants and medical students

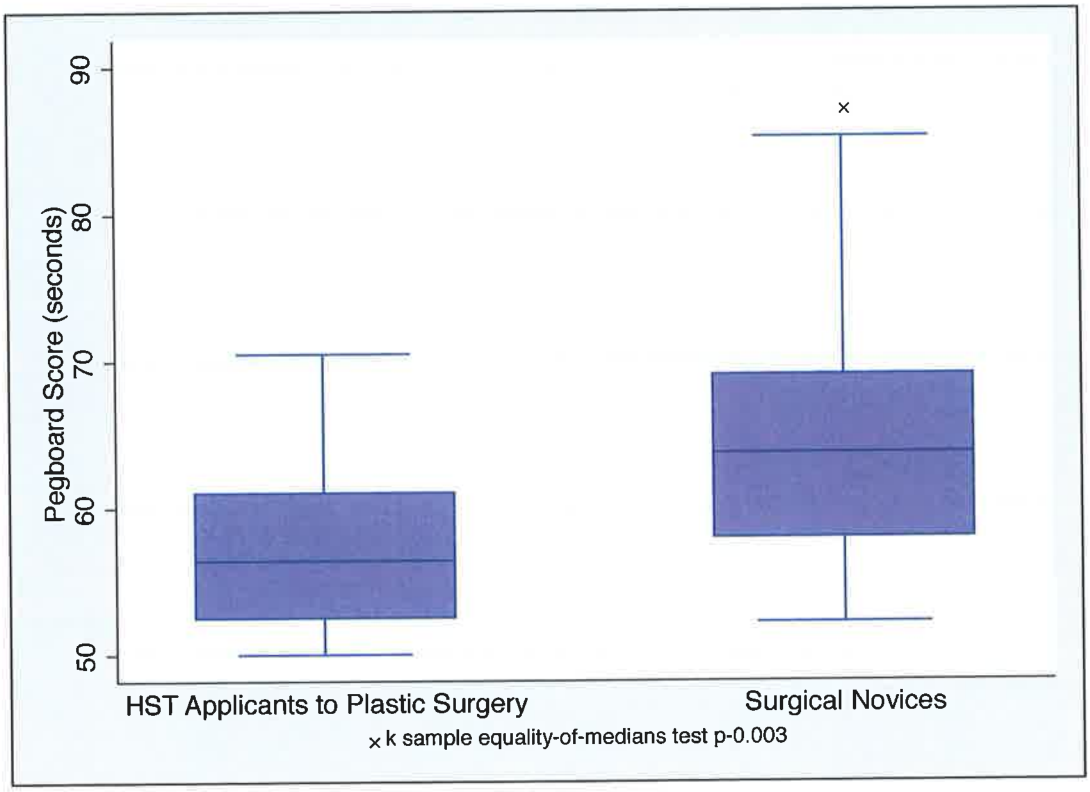


Figure 5.8 Psychomotor scores of general surgery HST applicants and medical students

5.4.5 Underperformance of HST's

A proportion of the plastic surgery HST applicants scored below the mean aptitude scores of the medical student population over the 6-year period. 11% scored below the mean score of the medical student population in visual spatial testing and 6% scored below the mean score of the medical student population in psychomotor and perceptual testing.

In the general surgery HST group, 12% scored below the mean score of the surgical novice group in visual spatial and psychomotor aptitude. 9% of the general HST group scored below the mean depth perception score of the medical student group.

5.4.6 Successful Entry into Higher Surgical Training

The mean aptitude scores between the applicants who were successful and unsuccessful into HST were also compared. The results for the HST applicants to plastic surgery show that there was no statistical difference between these two groups in any of the three areas of aptitude (Table 5.2). For the general surgery HST applicants, there was no difference between the visual spatial and the psychomotor aptitude scores between the successful and unsuccessful groups, however there was a significant difference in depth perception ability between the two groups.

Table 5.2. Aptitude scores of HST applicants

	HST Applicants (General Surgery)		N=113	HST Applicants (Plastic Surgery)		N=41
	Successful	Unsuccessful	p value	Successful	Unsuccessful	p value
Visual Spatial						
Median	65.7	64	0.3	58.8	61.3	0.44
Range	(25-94)	(33-100)		(38-84)	(16-82)	
Depth Perception						
Median	94	91	0.016	90	93	0.52
Range	(78-96)	(53-98)		(61-97)	(84-98)	
Psychomotor						
Median	58.2	61.6	0.09	60.9	58.3	0.3
Range	(44-86)	(48-70)		(54-69)	(52-61)	

5.4 Discussion

A candidates innate ability has been used in the assessment of applicants for technology based skills from fighter pilot to production line workers. This is considered vital in identifying attributes, which would predict good performance in these careers. Surgical trainee selection should be no different. Training a doctor to expertise in surgery is furthermore a long and expensive process and selection processes should therefore be robust and scientific.

The aim of good training program is to ensure all trainees excel in their chosen field. Trainees want to be the best that they can be and work hard to that end. It is therefore very sad to see a bright young trainee struggle in in their training years and potentially have difficulties in independent practice subsequently. It is therefore vitally important that the selection process into HST be as thorough and robust as possible. There is a duty to the candidates that the selection process is the one most likely to select candidates that will be successful. Finally we need to be mindful that the selection of candidates who go on to underperform in independent practice may result in legacy issues for those involved in the selection process. In our continual pursuit of the best selection methods we wanted to analyse the degree of self-selection of candidates in to plastic surgery.

From the data analysed in this study, it is apparent that HST applicants possess fundamental abilities considered important in the field of surgery. It is also clear that they have consistently higher aptitude scores than those of candidates who wish to pursue a career in surgery. This is reassuring and to be encouraged.

Although there appears to be a self-selection process of candidates with high innate ability into surgery, not all of the candidates had high performance scores. There were a small percentage (7.6% for plastic surgery and 11.6% for general surgery) of applicants over the six-year period that underperformed in relation to the medical student population.

This shows us that although the majority of candidates (greater than 90%)

naturally possess good fundamental ability appropriate to becoming a competent surgeon, there is a group who underscored in aptitude assessment. This is an important point to consider as this minority group has progressed through basic surgical training, which is a significant portion of postgraduate training without being flagged as potential weak candidates.

For such a candidate who has already spent several years of training in a surgical environment, to accept that they may not have the appropriate aptitudes to be a surgeon is difficult if not impossible. It is therefore reasonable to propose that testing should be performed on middle grade medical students. This group would likely be more amenable to advice based on their abilities, as they would not have invested time, energy and emotion in pursuing a specific field.

Our data stimulates the debate on the use of the assessment of innate ability as part of any training programme. At this stage it is unclear whether the group who progressed through BST and onto HST despite low aptitudes scores were able to compensate by putting in additional hours or if they could potentially be struggling at an advanced stage in their career pathway.

Chapter 6

Zone Calculation as a Tool for Assessing Performance Outcome in Laparoscopic Suturing

6.1 Introduction

Accurate performance assessment of surgical trainees is an essential component of our surgical training pathway and is fundamental for proficiency-based training. Laparoscopic performance is objectively measured by metrics produced by surgical simulators. The ProMIS III® simulator gives measurements of path length, smoothness of movement and time to completion of a laparoscopic task or procedure. Not only is it important to assess performance objectively but also to provide metrics that are meaningful and informative to the trainee.

Time is a simple calculation of the length of taken to perform the laparoscopic task or procedure. Path length is a measure of the distance travelled by an instrument during a procedure. Its measured by the simulator in millimeters (mm). Smoothness, which can also be called economy of movement, is a comparative score and has no units. It is essentially a measurement of any sudden changes in direction or acceleration of the laparoscopic instrument. Sharp turns create a high value while smooth movement creates a low value. All three of these metrics as produced by the simulator are calculated as cost functions, in which a lower value indicates a superior performance. These metrics have shown to be good indicators of task progression in laparoscopic surgery and studies have demonstrated their validity. It would be ideal if metrics were also informative and meaningful to the user performing the laparoscopic task on the simulator.

In order to perform laparoscopic suturing, one must be able to perform this complex task within a confined space making it one of the more advanced laparoscopic tasks. Smoothness which is essentially the recorded path length compared with a calculated optimal path length gives an indication of the global performance, but does not provide information on the performance of specific tasks. Similarly length travelled by the user with the laparoscopic needle holders does not provide feedback on how well an intracorporeal knot was tied for example.

Feedback can be defined as the return of performance-related information to the trainee. It is termed as extrinsic, when it is provided by an external source (Ende, 1983). Extrinsic feedback has to be meaningful and informative, which is traditionally in the form of expert feedback. In the absence of this, metrics can provide objective feedback. Automated feedback has been demonstrated to have similar efficacy to live expert feedback (Snyder et al., 2011, Xeroulis et al., 2007). However, these metrics need to be meaningful to the trainee.

In order for the simulator to calculate metrics, it records the 3-D position angle and time-stamp for the laparoscopic instrument 15 times per second. This is stored in one file called a 'pointstream' which is a complete history of that instrument for any given performance on the simulator. All metrics are calculated by analysing this file. Time is simply a subtraction of the first time-stamp from the last. Path length is the addition of all the distance points (using the distance between each 3-D point recorded). In order to calculate smoothness, a curve analysis is performed. In our chapter, we decided to take this point stream and perform our own interpretation of the data by creating a new metric to measure laparoscopic suturing.

We attempted the develop new metrics called “in-zone” score and “out-zone” score in order to provide the user as to the percentage of time spent with their instruments in specific areas during a laparoscopic suturing task. We felt that this would be more informative than motion efficiency and path travelled.

Previous efforts have been made to validate a meaningful assessment method for laparoscopic suturing in a study by Botden et al. They devised a graphic dome, which was overlaid on the suturing pad and laparoscopic instruments during performance assessment. They calculated the time spent within the dome when throwing the thread around the needle holder, the time spent outside the dome when tying the knot and the strength of the knot. One of the limitations of this study was that the participants regarded the dome as a hindrance and obstructed their vision of the instruments and suturing thread while performing their task.

In an attempt to design a scoring method to accurately assess laparoscopic suturing, we designed a “zone” system, which allowed us to calculate the exact area within a three-dimensional space, in which a surgeon operates. The primary aim of this study was to assess if this new “zone metric” was a valid method of assessing laparoscopic suturing.

6.2 Objectives

6.2.1 Hypothesis Underlying the Objectives

It was hypothesised that measuring the exact area wherein one is operating during a laparoscopic suturing task would provide a valid assessemnt tool of laparoscopic suturing. As one must be able to perform this complex task within a confined space with precision to avoid damage to the surrounding structures, calculation of the area where one is operating was thought to be of more relevance than the length travelled by the instrument.

Our aim was to validate this newly created "zone" metric by correlating it to both subjective performance scores as well as the traditional metrics of path length, smoothness and time.

6.2.2 Detailed Objectives

Objective 1. To provide concurrent validity for the novel "zone" metrics.

We attempted to establish concurrent validity, by correlating the zone metrics to the subjective blinded observer scores (FLS and OSATS). We aimed to further validate the new metric scores by correlating the zone metrics to the traditional metrics of path length, smoothness and time. These correlations were done using Spearman's Rho.

Objective 2. To provide construct validity for this novel objective assessment method for laparoscopic suturing.

We aimed to establish construct validity by comparing the zone metric scores of the laparoscopic suturing performances between three groups with different levels of laparoscopic experience.

6.3 Materials and Methods

6.3.1 Participant Recruitment

As outlined in chapter 2, participants were recruited to take part in this study on a voluntary basis. An email was distributed to all medical students in RCSI with details of the research project. An email was also send out to the first and second year BST's calling for volunteers to take part in this study. As laparoscopic suturing is considered one of the most complex laparoscopic tasks, both first and second years were eligible for this experiment. If any of the trainees however had performed laparoscopic suturing as part a laparoscopic procedure, they were excluded from the study. Five laparoscopic experts were also recruited in order to set benchmark proficiency levels on the ProMIS simulator. Each of these experts had performed over 150 complex laparoscopic procedures.

All participants who were selected were asked to sign a consent form allowing all data collected to be used for research purposes. It was made clear to all the subjects that the data was stored and presented in an anonymous format.

6.3.2 Participant Demographics

Ten medical students were recruited to participate in this research study. They all had similar baseline ability. Ten basic surgical trainees were recruited to take part in the study. They all had similar operative ability. The inclusion and exclusion criteria for the study participants have been previously outlined in chapter 2, materials and methods.

6.3.3 Zone Calculation

A software program was developed to analyse the point stream data as previously described. This allowed for the creation of an additional metric to track the location of the instruments in a three dimensional space during the task performance.

The operating field was divided into 10 virtual zones, each zone was 1 cm and its distance was measured from the base of the operating field. Figure 6.1 depicts how the operating space was divided into virtual zones. We decided that between 0 and 6 cm from the base of the suturing pad would be the "in-zone" which included the first 6 inner zones. The "out-zone" would be 6 and 10 cm from the base of the suturing pad and included the 4 outer zones (Figure 6.2).

The dimensions of the in-zone (0-6 cm) and out-zone (6-10 cm) areas were derived from experienced laparoscopic surgeons during the suturing part of a laparoscopic rectopexy. The in-zone was the ideal space (6 cm distance from the base of the operative field), which the operator should stay within for the majority of time during the suturing task, allowing the operator to handle the instruments and throw the knots with ease. It is necessary to leave the in-zone and enter the out-zone space (beyond 6 cm) in order to secure and tighten the knot but this is only required for a short period during the procedure.

These new in and out-zone scores became our new metrics. In order to calculate the new zone metrics, the software program simply calculated the distance between the centre point of the operative field and the tip of the instrument. Depending on what range this given distance fell within, time spent in that zone was calculated which was measured as a "zone score".

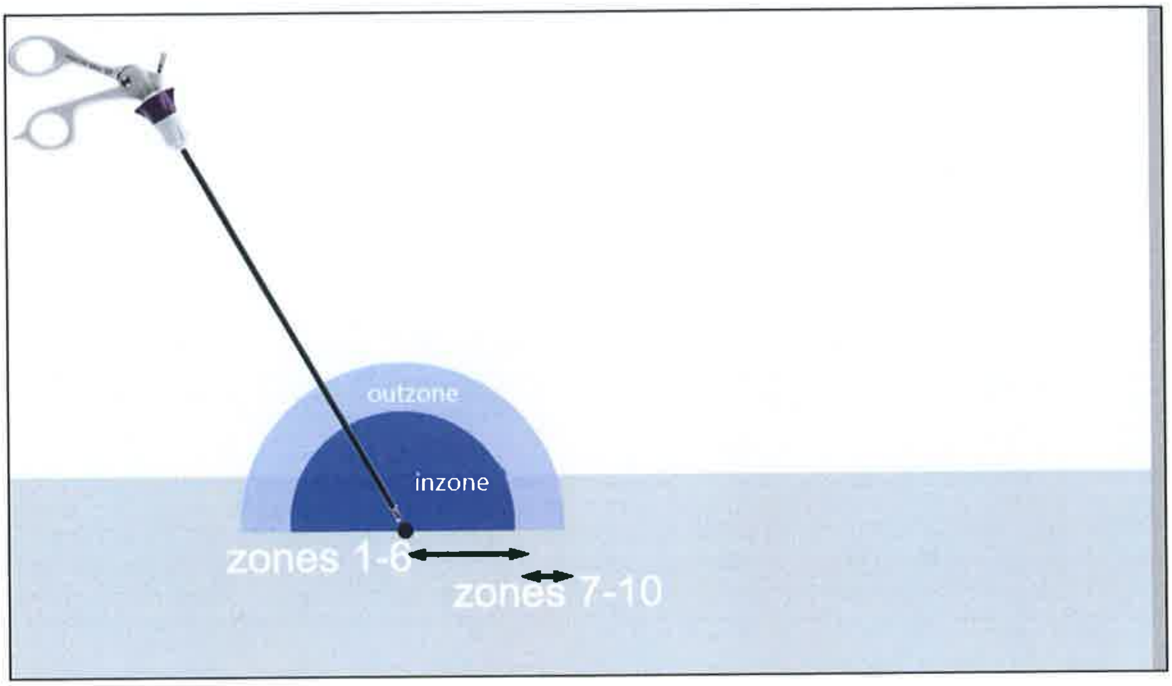


Figure 6.1 Graphical representations on the in-zone and out-zone areas

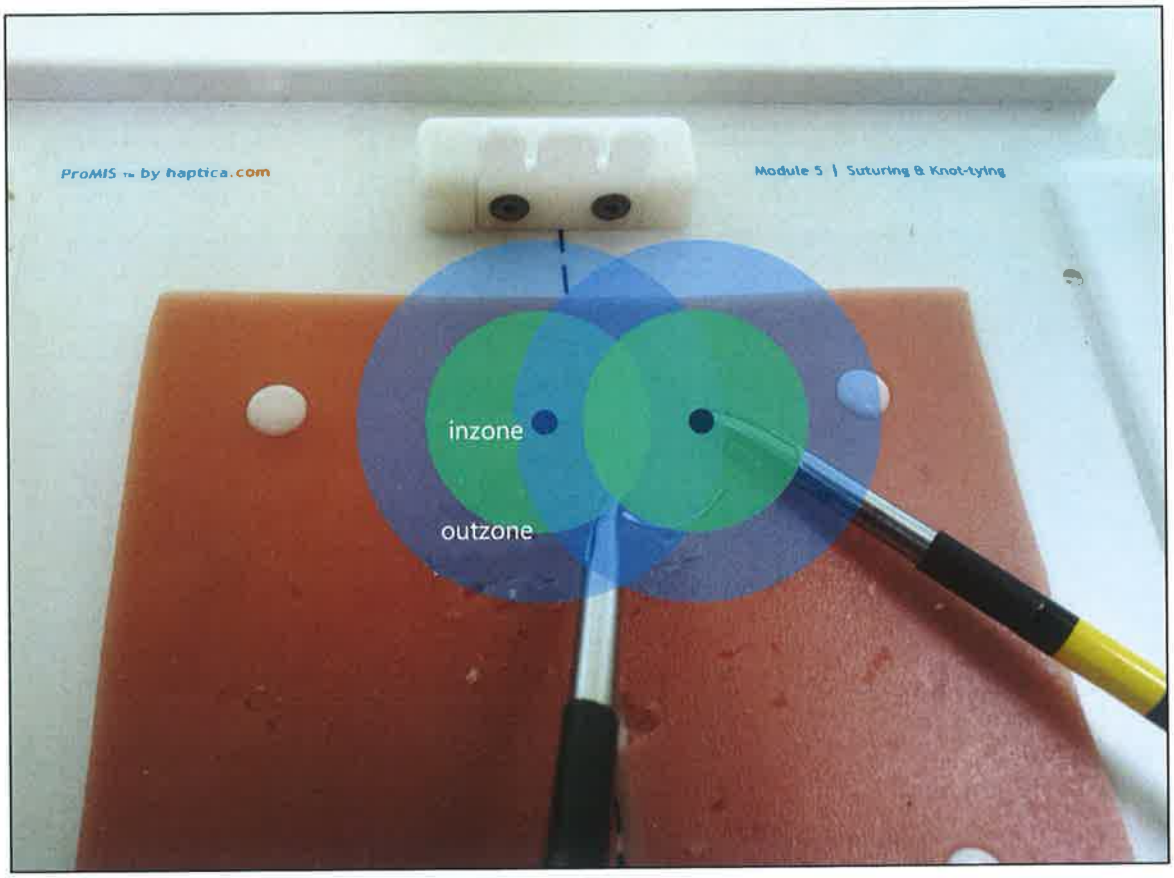


Figure 6.2 In-zone and out-zone areas in the laparoscopic suturing module

6.3.4 Surgical Skills Assessment

The ProMIS Simulator was used for assessment. The laparoscopic intracorporeal suturing and knot tying module was used for performance assessment. Synthetic suturing skin (The Chamberlain Group, Massachusetts®) was cut to size and a 10 x 12 cm piece was inserted into the ProMIS simulator tray. The laparoscopic suturing task was performed in a conventional way using 3/0 silk and two laparoscopic needle holders. Three knots were thrown in order to complete the task.

Prior to performance assessment each candidate received didactic teaching. Each candidate was sent a stepwise approach detailing how to perform the task and a video-link to a live recording of a specific part of a laparoscopic rectopexy before the experiment commenced. When the candidate attended the first session, a simulated laparoscopic suturing demonstration was conducted. They were allocated time to ask questions before they attempted a mandatory multiple-choice questionnaire to ensure complete comprehension of the procedure. Prior to their assessment, they had an opportunity to familiarise themselves with the testing equipment by completing a basic laparoscopic task. They were required to take a bite of skin using the laparoscopic needle holder and attached needle while stabilising the tissue with the second needle holder.

The virtual zones were not displayed on the screen for the candidates, as we were concerned that this would hinder their performance, by obstructing their vision of the instruments and suture material. Instead it was explained to that the objective of the study was to validate a new metric, which calculated the location where they were operating within the operative field. They were asked to try and stay within a 6 cm radius while performing the task apart from when they had to tie and secure the knot. In general they were told to operate close to the base of the field and not to wander with their instruments.

6.3.5 Performance Scores

The candidates performed the task under the supervision of a senior surgeon. If the subject needed guidance, instructions were given however at no stage during any of the perfromance asssesments did the senior surgeon take over the task.

In order to calculate the in and out-zone metrics, the software program analysed the point stream produced by the simulator (location of the instruments in the 3-D space) during the task performance and calculated the percentage time spent within the predefined in and out-zones. The traditional metrics of path length, economy of movement and time produced by the ProMIS simulator were also recorded.

Subjective scores were obtained by subsequent video analysis of the performances by two reviewers who were blinded to the experience level of the candidate. Objective Structured Assessment of Technical Skills (OSATS) and Fundamentals of Laparoscopic Surgery (FLS) scores were the subjective scales used to assess the suturing performances. Error score was calculated using is a list of pre-defined errors, which can be found in table 2.

6.3.6 Statistical Analysis

Data was analysed using Stata 12.0. Correlations between new zone scores and observer scores were performed using spearmans rho in order to determine concurrent validity. The three groups were compared using the kruskal-wallis equality-of-populations rank test in order to determine construct validity.

6.4 Results

6.4.1 Participant Demographics

28 subjects took part; 10 medical students, 10 basic surgical trainees and 5 expert surgeons. Demographics for the groups in terms of age, gender, handedness and videogame, sporting and musical ability are outlined in table 6.1. Prior operative experience for the ten BSTs is displayed in table 6.2. All expert surgeons had performed more than 150 complex laparoscopic procedures.

Table 6.1 Demographics of medical students in the laparoscopic zones group

	Medical Students (N = 10)
Age (years)	
Range	19-30
Mean	22.5
St. Dev.	4.5
Gender (%)	
Male	30
Female	70
Dominant Hand (%)	
Right	100
Left	0
Corrected Vision (%)	
Yes	50
No	50
Video Games (%)	
Yes (at least one hour/week)	20
No	80
Music (%)	
Yes (achieved distinction)	60
No	40
Sport (%)	
Yes (intercollegiate level)	50
No	50

Table 6.2 Demographics of BST's in the laparoscopic zones group

	BST's (N = 10)
Age (years)	
Range	25-32
Mean	27.1
St. Dev.	2.3
Gender (%)	
Male	50
Female	50
Dominant Hand (%)	
Right	100
Left	
Corrected Vision (%)	
Yes	70
No	30
Video Games (%)	
Yes (at least one hr/week)	60
No	40
Music (%)	
Yes (achieved distinction)	50
No	50
Sport (%)	
Yes (intercollegiate level)	60
No	40
Operations Performed Unsupervised (Mean Number)	
Excision of Lesion	30
OGD	3
Laparoscopic Appendicectomy	1.5

6.4.2 Analysis of Operative Time Spent within Zones

Upon reviewing the results of the “zone scores” of each group, it was clear that the expert group spent the majority of their time within zone 3 to 6. The BST group also showed a trend toward spending the majority of their operative time in zone 3 to 6 but did spend more time in the outzones of 7 and 8 than the expert group. The novice did not show any specific pattern in which zones they spent the majority of their operative time during the laparoscopic suturing task (Figure 6.3 and 6.4).

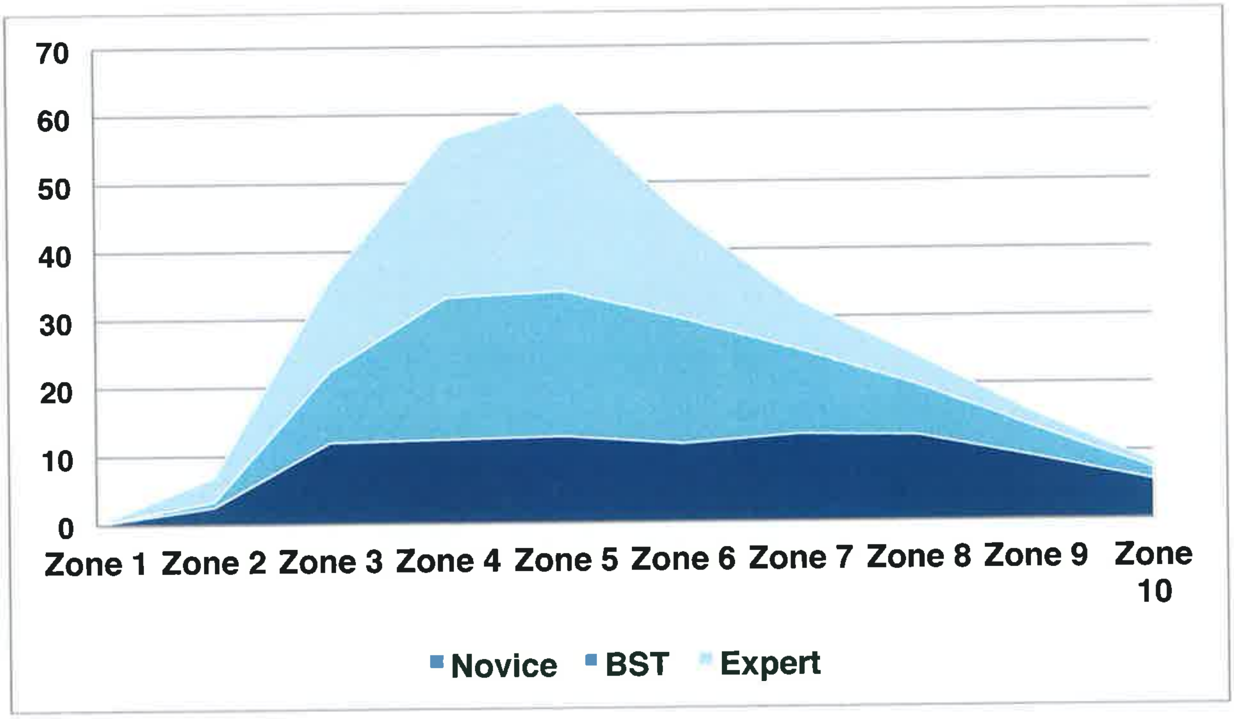


Figure 6.3 Percentage of time spent in right zones

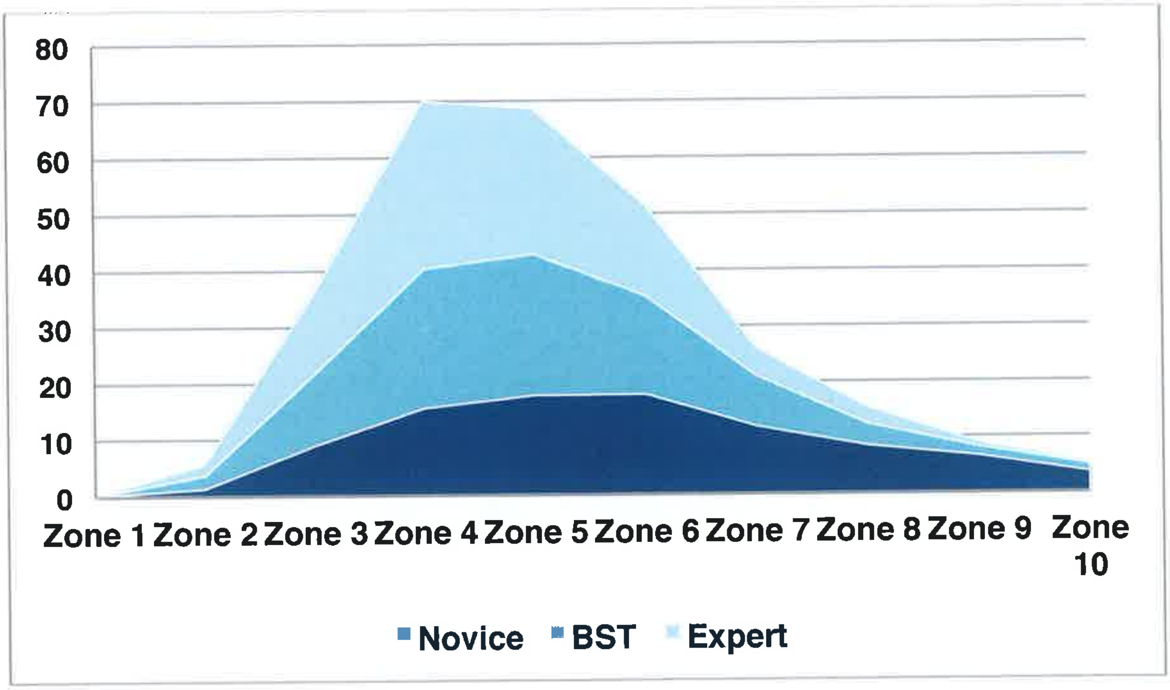


Figure 6.4 Percentage of time spent in left zones

6.4.3 Comparison of "In-zone" Scores

There was a significant difference in the average in-zone scores between all three experience groups (p=0.0001). The average expert left in-zone score was 83% and right in-zone score was 88%. This implied that one should be spending 83% and 88% of time with the left and right needle holders respectively within a 6 cm distance from the base of the field. The BSTs spent 69% and 72% and the medical students spent 50% and 49% within the left and right in-zones respectively during their first attempt at the laparoscopic suturing task. These results are displayed in Figure 6.5.

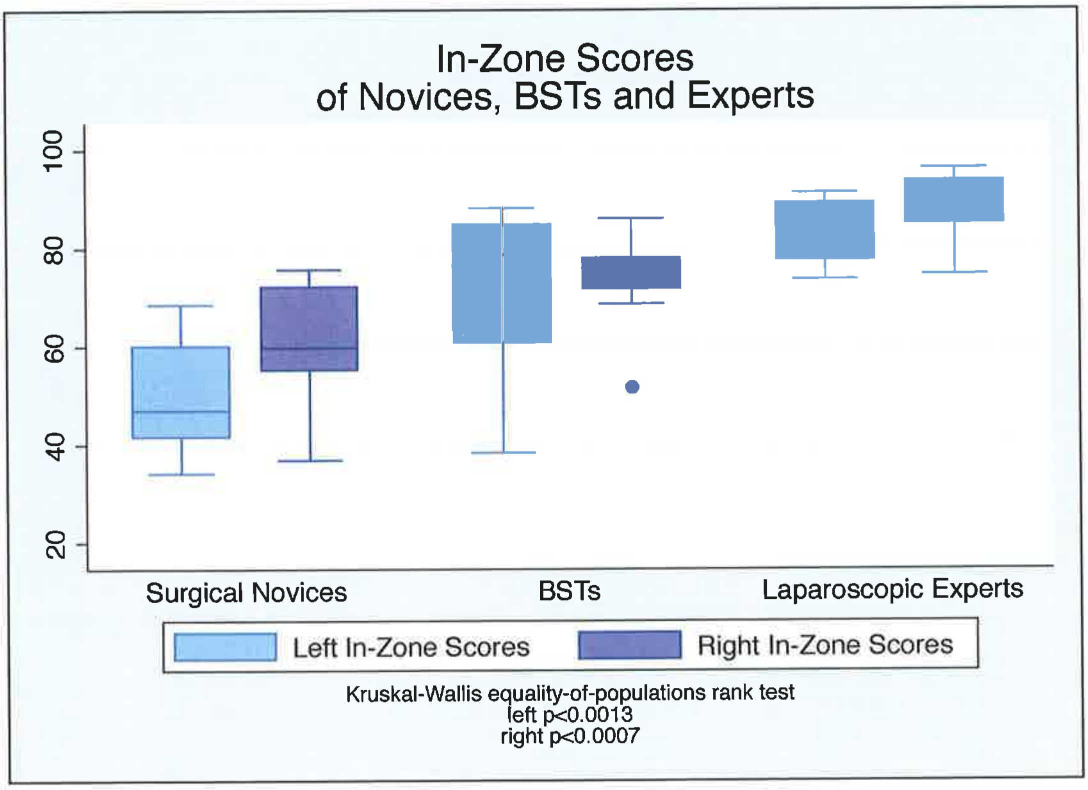


Figure 6.5 In-zone scores of all groups

6.4.4 Comparison of "Out-zone" Scores

There was also a significant difference in the average out-zone scores between all three groups (p=0.0001). The average expert left out-zone score was 16% and right out-zone score was 12%. This was the maximum amount of time that should be spent outside the 6 cm inzone which was required during tightening and securing the knot. The BSTs spent 31% and 24% of time and the medical students spent 61% and 38% of time within these unsafe areas on the left and right respectively during their first attempt at the laparoscopic suturing task. These results are displayed in Figure 6.6.

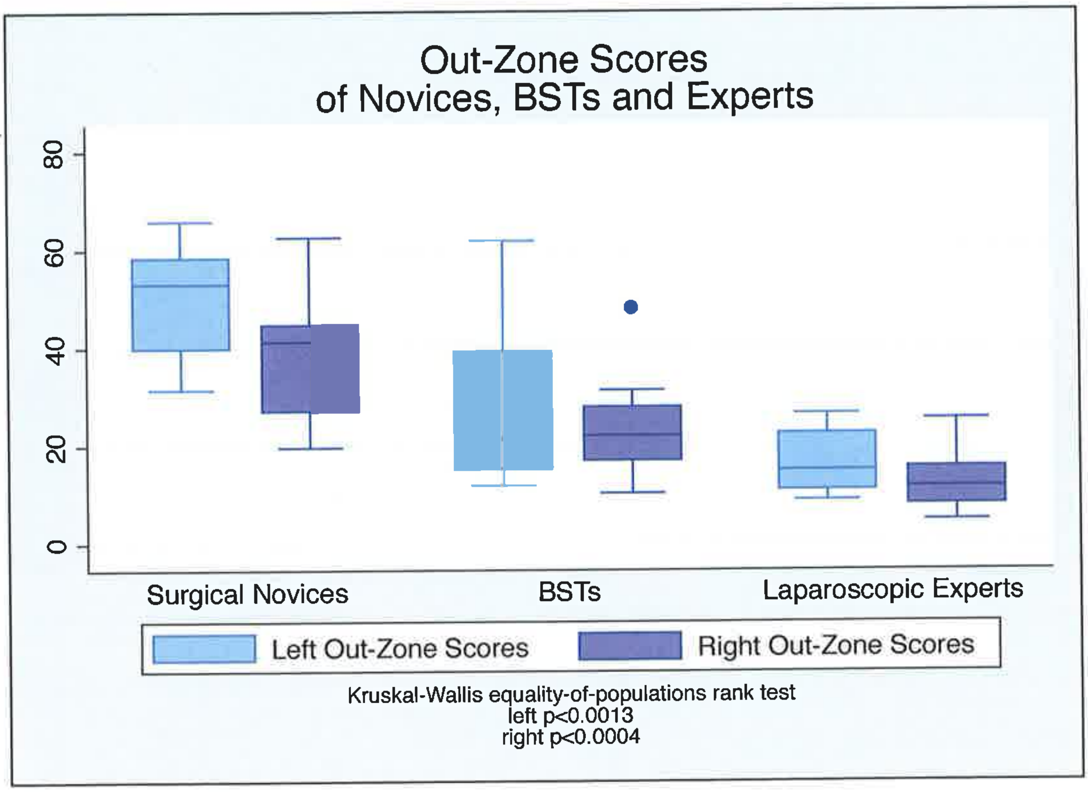


Figure 6.6 Outzone scores of all groups

6.4.5 Comparison of Subjective Scores

The Fundamentals of Laparoscopic Surgery evaluation was scored out of 30. The Observed Structured Assessment of Technical Skills evaluation was scored out 25, as knowledge of instruments and use of assistants were excluded as they did not apply in this case. There was a significant difference between the FLS and OSATS scores in all three groups which are shown in Figure 6.7. The average FLS scores were 28 (out of 30) in the expert group compared to 13 in the BST group and 9 in the novice group. The average OSATS scores were 24 (out of 25) in the expert group compared to 13 in the BST group and 6 in the novice group.

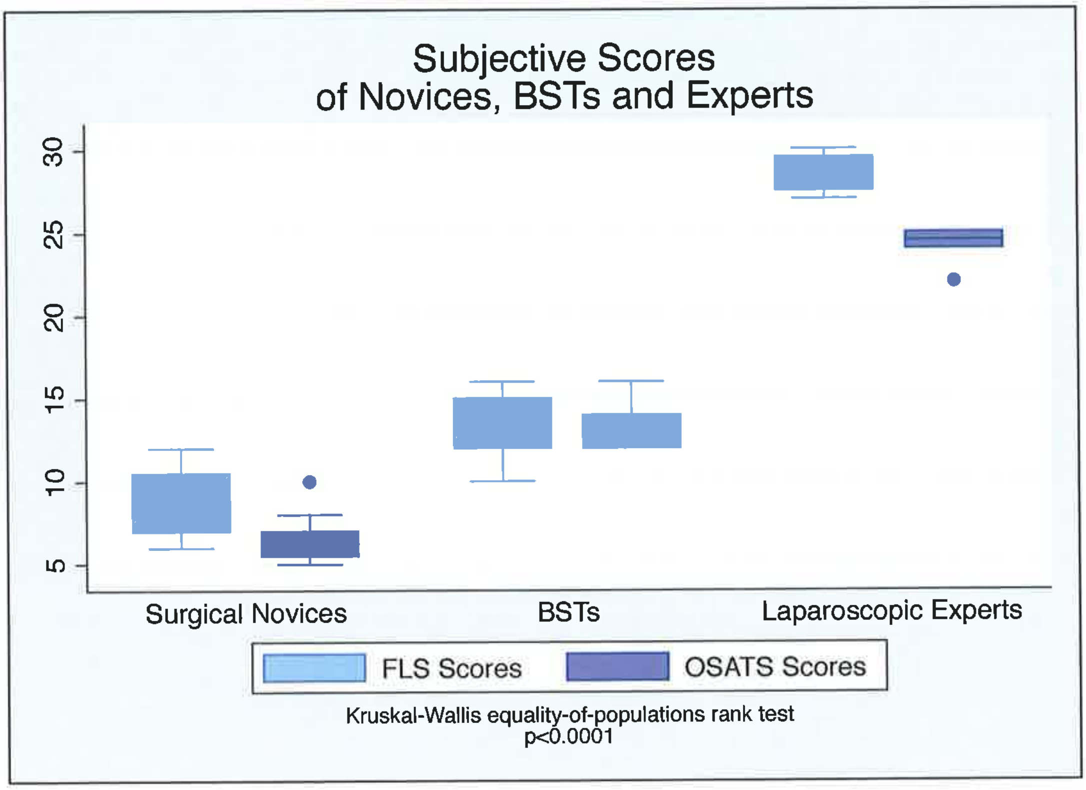


Figure 6.7 Subjective scores of all groups

6.4.6 Correlations between the Zone Scores and Subjective Performance Scores

In order to establish concurrent validity, the new zone metrics were compared with subjective scores (OSATS and FLS). Reviewers blinded to the experience level of the candidates established these subjective scores. The new zone metrics scores correlated significantly with both the subjective blinded-observer scores of OSATS and FLS. The correlation results are displayed in Table 6.3

Table 6.3 Correlation between zone scores and subjective scores

N = 25	FLS		OSATS	
Left In-zone	$r = 0.864$	$p < 0.0001$	$r = 0.88$	$p < 0.0001$
Left Out-zone	$r = 0.864$	$p < 0.0001$	$r = 0.88$	$p < 0.0001$
Right In-zone	$r = 0.811$	$p < 0.0001$	$r = 0.822$	$p < 0.0001$
Right Out-zone	$r = 0.782$	$p = 0.0001$	$r = 0.795$	$p = 0.0001$

6.4.7 Correlations between the New Zone Scores and Traditional Metric Scores

To further illustrate concurrent validity, the new zone scores were correlated to the traditional method of measuring laparoscopic performance on the ProMIS simulator (path length, smoothness and time). The new zone metric scores correlated significantly with the traditional metrics of path length, time and smoothness. The correlation results are displayed in Table 6.4.

Table 6.4 Correlation between zone scores and traditional metrics

N = 25	Time		Path Length		Smoothness	
Left In-zone	*r*=0.793	p=0.0001	*r*=0.801	p=0.0001	*r*=0.801	p=0.0001
Left Out-zone	*r*=0.793	p=0.0001	*r*=0.801	p=0.0001	*r*=0.801	p=0.0001
Right In-zone	*r*=0.735	p=0.0005	*r*=0.775	p=0.0002	*r*=0.754	p=0.0003
Right Out-zone	*r*=0.718	p=0.0008	*r*=0.763	p=0.0002	*r*=0.742	p=0.004

6.4.8 Validation of a New Objective "Zone" Metric for Laparoscopic Suturing

In order to show construct validity, the average perfomances scores between the three groups tested were compared across all parameters. The results show that surgical trainees performed significantly better than medical students. The experts performed significantly better than both the medical students and the surgical trainees. The results of the mean scores for the new zone metrics are displayed in Table 6.5.

Table 6.5 Metric and subjective scores in all three groups

	Medical Students N = 10	BST's N = 10	Experts N = 5	p value (Kruskal-wallis test)
Left In-zone	50.5	69.8	83.5	0.0013
Left Out-zone	61.2	30.6	16.5	0.0013
Right In-zone	49.5	72.7	88.3	0.0007
Right Out-zone	38	23.9	12.5	0.0004
FLS	8.9	13.3	28.5	0.0001
OSATS	6.5	13.6	24.2	0.0001

6.5 Discussion

In this chapter, we calculated in-zone and out-zone scores for a laparoscopic suturing task across three groups of candidates with a wide disparity of surgical expertise. Our aim was to established concurrent validity by comparing the zone scores to calculated subjective observer scores. We also aimed to demonstrate construct validity by comparing the new performances scores (zone metrics) between the three groups of candidates with differing levels of expertise.

Our hope in developing these "zones" as an assessment tool was to try and provide a meaningful measurement of laparoscopic suturing performance. The zone classification was devised based on the performance scores of experts using a software program to analyse the point stream data produced by the simulator.

Traditional parameters such as time, path length and smoothness are excellent measures for assessing task progression for laparoscopic procedures (Mason et al., 2013). However they provide information to the user about global performance rather than performance outcome in a specific laparoscopic task. Therefore by creating a three-dimensional space that the candidate has to stay inside while throwing the suturing thread around the needle-holder allows accurate measurement of the task. This space was imagined as zones based on the movements travelled by experienced laparoscopic surgeons while performing a suturing task. This is also the ideal space within which to stay during suturing therefore the candidate learns to suture within a confined space.

Providing a novel assessment tool for laparoscopic suturing performance could potentially be used in formal assessments as it has been validated against expert derived subjective scores as well as our known validated traditional metrics. By developing an objective validated metric of laparoscopic suturing which is both meaningful and informative for the trainee, we have devised a novel method of assessment. In order for simulators to train surgeons effectively, it is a necessary that there are appropriate feedback mechanisms.

The results of our study confirm that calculation of percentage time spent within specific zones is a valid method of assessing laparoscopic suturing. The new zone scores correlated with the traditional metrics as well as subjective observer scores thereby supporting concurrent validity. The performance scores of our zone assessment method showed significant differences between all three experience groups and thereby supporting construct validity.

In conclusion, the new metric is a valid tool for assessing laparoscopic suturing objectively. It potentially could be incorporated into the development of technical safe surgical practice using VR zone based training. This new zone metric has shown to have both concurrent and construct validity.

Chapter 7

Creation of a Meaningful Objective Scoring System Using Metrics Produced by the ProMIS Simulator

7.1 Introduction

Structured training is vital for successful surgical education programs. Surgical trainees need performance goals set by experts. They also need appropriate validated methods of assessing their progress during the surgical training pathway. Moreover is it important that mentors and surgical training bodies have a standardised approach to assessing trainees throughout their training as surgery is a technically demanding field which requires very high operative standards.

In order to use metric scores to objectively measure the progress of surgical trainees, appropriate validated scoring systems must be developed. The ProMIS simulator is very useful for providing objective feedback. The candidate carries out a laparoscopic basic skill task on the ProMIS simulator, which tracks the instruments in 3-D space, analyses the instrument movement and creates metrics that are based on that analysis. These metrics are time, path length and smoothness. However it is undetermined how these metrics can provide a score in relation to a known standard such as expert scores. Further to this, the simulator records several metrics which are all recorded in different units making it difficult to use them in a formal assessment process.

A common criticism of simulator-derived metrics is the difficulty experienced by trainees in understanding their metrics scores. This is frequently encountered during assessment of trainees on laparoscopic simulators. There is difficulty in interpreting the metrics produced by the simulator and it is unclear as to what percentage scores to assign the trainees. This must be considered in order to engage in objective continuous assessment as part of the surgical training pathway.

The ProMIS gives a report of three metrics. Time is measured in seconds; path length, which is measured in millimeters and smoothness, which is purely a comparative score, has no units. It is difficult for the trainees to understand their score and to gauge how well they have performed in comparison to an expert score.

Previous studies have attempted to create scoring systems within specific task trainers. The computer enhanced laparoscopic training system (CELTS) was developed by the Centre for the Integration of Medicine and Innovative Technology CIMIT and Harvard Medical School. They used a box trainer with a computer interface to form a task-independent scoring system against expert benchmark levels. Expert scores were calculated for suturing, peg transfer and knot tying using time, path length, smoothness, and depth perception as metrics. The user score was then compared with an expert score which led to the development of a standardised scoring system.

7.2 Objectives

7.2.1 Hypothesis Underlying the Objectives

The aim of this study was to develop an objective scoring system which would provide meaningful user scores and appropriate feedback of performance in relation to experts. We wanted to create a single score that would represent the existing validated objective scores as a single percent score.

7.2.2 Detailed Objectives

Objective 1. To develop an meaningful objective unified score using combined metrics produced by the ProMIS.

We aimed to standardise the assessment of surgical trainees by creating a unified scoring system based on the three objective metrics prodiced by the ProMIS. This "unified" score would be a percentage score which would be meaningful to the trainee using the lapaoscopic simulator.

Objective 2. To establish a method of comparing performance scores of surgical trainees

We aimed to create a formula which would compare a trainees performance score to the performance scores of a large number of surgical trainees over a three year period.

Objective 3. To establish a method of comparing the trainees laparoscopic scores to experts scores.

As part of a proficiency based training programme, trainees strive to achieve the top performance goals in a simulated training environment. Using the new formula, we also aimed express the trainees performance scores in relation to experts proficiency scores.

7.3 Materials and Methods

7.3.1 Participant Recruitment

In Ireland, surgical training consists of two phases. Basic Surgical Training (BST), which is followed by Higher Surgical Training (HST). Selection for HST is competitive and only the top performing graduates (approximately 30%) from our BST programme's progress to HST.

In order to progress within BST and equally to progress onto HST, part of the assessment process includes a technical skills assessment. Trainees in the second year of BST and shortlisted HST candidates were included in this study. The Research Ethics Committee of RCSI granted ethical approval. All BST and shortlisted HST candidates provided written informed consent to supply data collected as part of the assessment process analysed and used for research purposes.

7.3.2 Participant Demographics

A database of metric scores was maintained over a 3-year period (2011-2013) for 209 trainees.

7.3.3 Setting Proficiency Levels

5 laparoscopic experts was also recruited in order to set benchmark proficiency levels on the ProMIS simulator. Each of these experts had performed a minimum of 300 laparoscopic procedures.

7.3.4 Surgical Skills Assessment

For the assessment, each participant carried out 5 laparoscopic tasks. They performed three tasks of object positioning and two tasks of sharp dissection.

There are various levels in these modules with increasing complexity. The 'object positioning' module requires the operator to move beads from one pot to another on a pre-designed tray in the laparoscopic simulator. The 'sharp dissection' module requires the operator to cut a straight line and then cut out a triangle on a fixed glove. The simulator verbally instructs the operator at each step.

Upon completion of each procedure the simulator provides a summary report of the metric scores: path length, smoothness and time. This gives a total of 15 individual scores per candidate all measured with different units. Each performance recorded was also assessed using the OSATS scoring system. The OSATs evaluation was scored out of 25 as “knowledge of instruments” and “use of assistants” were excluded as they did not apply for these ltasks.

7.3.5 Creation of Unified Score

To establish an expert performance baseline database for each of the two tasks, a group of surgeons who are considered to be experts in laparoscopy completed each of the tasks.

In order to compare this expert performance to any subsequent performance by a trainee, we developing a formula which would allow us to assign a standardised overall score from 0 to 100 for the trainee for any given laparoscopic performance of these tasks on the ProMIS simulator. This method scored the trainees in relation to each other essentially as the range used was a wide range of performance scores of surgical trainees over a three-year period.

The formula was derived on the basis that each individual score should be expressed mathematically in a universal unit as every metric (path length, smoothness and time) had differing units. The obvious choice here is to use percentages, as it is the universally accepted method of scoring any exam or performance.

In order to determine a range from which we could design percentage scores, we took the worst performance of the group and made that number equal to zero, and we took the best performance and made that equal to 100. The expert scores were used as a marker for the best score, which was equal to 100. In trying to unify the score, all the scores in between the min and the max were then assigned a percentage based on this range according to the formula, which is detailed in Table 7.1 and displayed in Figure 7.1.

The formula as it is written here, is a mathematical way of expressing our set of numbers. From this set we identified the range using the largest and smallest numbers. The range was divided by 100, which gave us a figure that converted the original numbers into percentages. The software used for the calculations was Microsoft Excel. Using the inbuilt formulas, we ran a MAX and MIN on the column that produced the top and bottom numbers from the data set that fed into our formula.

The ProMIS Simulator automatically creates the metrics for each task (path length, smoothness and time) and these were manually imported into the data sheet. This Excel sheet provided the source data and working spreadsheet to create the new column that contained the unified score along with the calculations used to produce this unified score.

Table 7.1 Details of the conversion of metrics to a percentage score

$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$	Where a_1, a_2... are raw scores from the simulator
C	The number of individual tasks carried out by each candidate
X=min (a)	Lowest score
Y=max (a)	Highest score
$Z=100/((a_n - x) * y)$	This metrics expressed as a percent. By subtracting the result from 100, it turns a low score into a high percent score

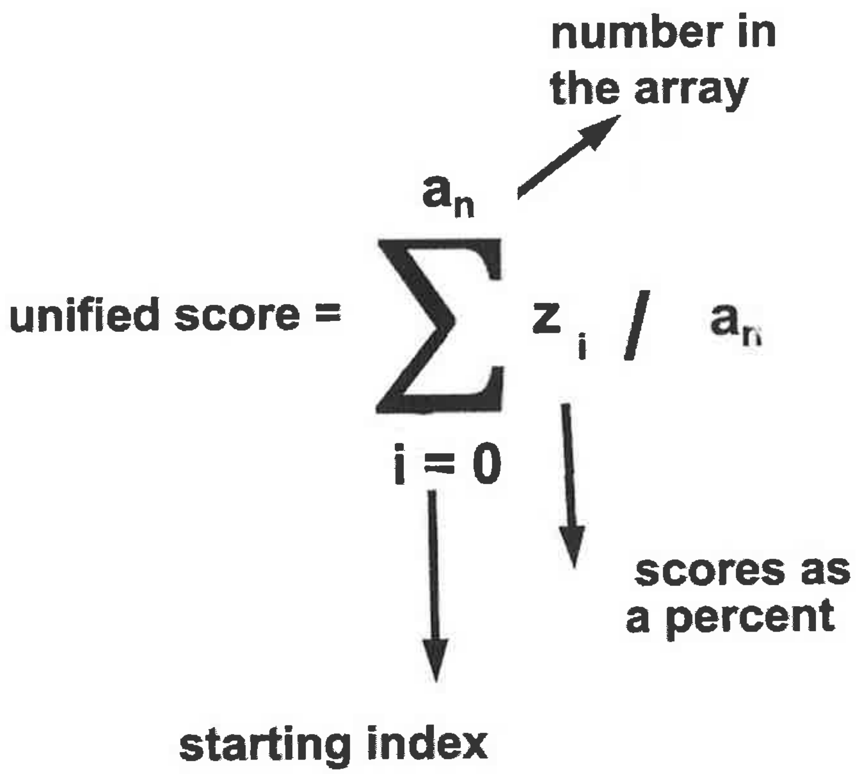


Figure 7.1 Formula developed to create a single user score

A similar approach to establishing a standard to compare trainee and expert scores performance on a laparoscopic simulator was developed in Harvard using the CELTS trainer. They utilised a z-score statistic in assign individual scores to trainees. Our method focuses on collating the metrics into one unified score, which is meaningful to the user. By defining 100% as an expert score, we created a process that allowed trainees to be scored in relation to an expert.

This formula was specific to the laparoscopic tasks of object positioning and sharp dissection (as previously detailed) on the ProMIS simulator but could be easily developed for any task on any trainer.

7.3.6 Statistical Analysis

In order to validate the newly created single score, these single user scores were correlated to both the objective metrics produced by the simulator and the subjective OSATS scores as recorded by the assessor. The formula was validated by correlating the percentage scores (calculated using the formula) with percentile ranks of the same data set. Correlations were performed using spearmans rho. Data was analysed using Stata 12.0.

7.4 Results

7.4.1 Participant Demographics

Of the 209 candidates, 164 were BST's and 45 were HST's and they were aged between 25-36.

7.4.2 Correlation of Unified Scores and Subjective Scores

The overall co-efficient correlation was $r = 0.46$ ($p<0.0001$) for the new unified scores compared with the subjective scores. Figure 7.2 demonstrates a scatterplot of these scores. When this was analysed for each group (experts, HST and BST's), the new unified scores correlated significantly ($p=0.001$) with the subjective scores (OSATS). The lowest correlation score was for the HST group. Breakdown of scores can be seen in Table 7.2.

Table 7.2 Correlation of new unified score and subjective scores

	Subjective Scores	
New Unified Score		
Experts	$r = 0.895$	$p < 0.0001$
HSTs	$r = 0.353$	$p = 0.001$
BST's	$r = 0.582$	$p < 0.0001$

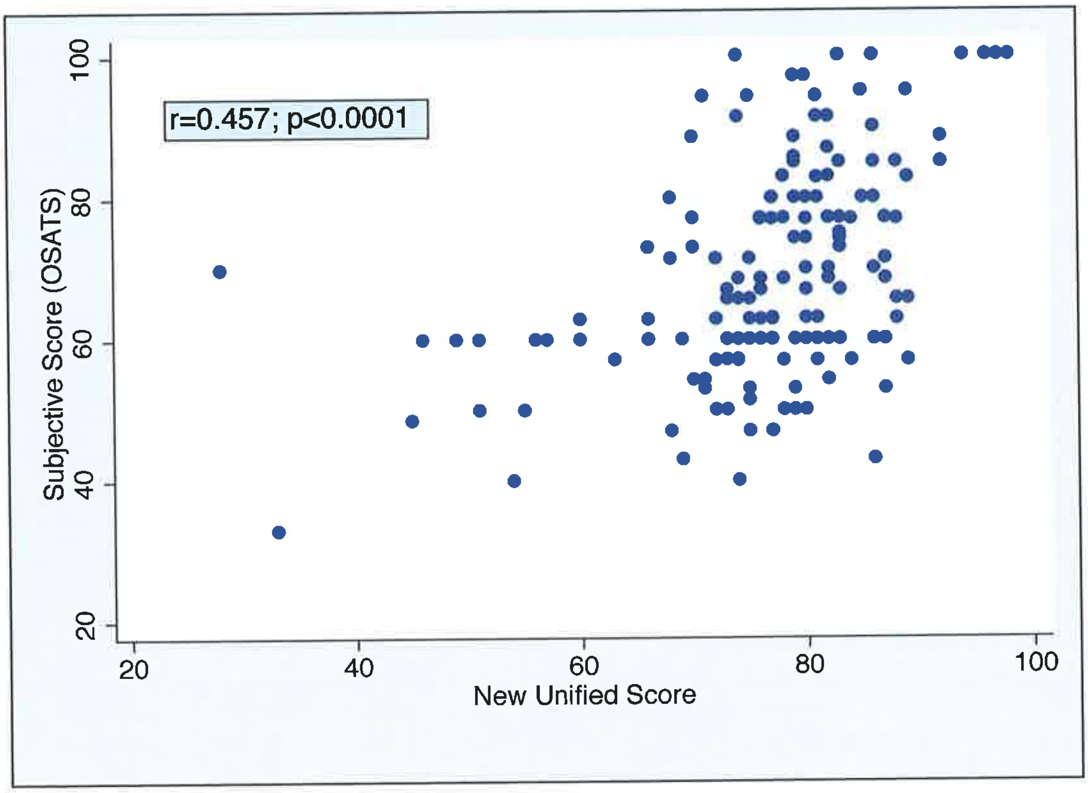


Figure 7.2 Scatterplot displaying correlation between the new unified scores and the subjective OSATS scores

7.4.3 Correlation of Unified Scores and Objective Metrics

The overall co-efficient correlation was $r = 0.84$ (p<0.0001) for the new unified scores compared with the three traditional objective metrics (time, path length and smoothness). Figure 7.3 displays this correlation. When sub-group analysis was performed, the co-efficient values remained statistically significant (p<0.001). A breakdown of these results can be seen in Table 7.3

Table 7.3 Correlation of new unified score and objective metrics

	Objective Metrics	
New Unified Score		
Experts	$r = 0.895$	$p < 0.0001$
HST's	$r = 0.834$	$p < 0.0001$
BST's	$r = 0.663$	$p < 0.0001$

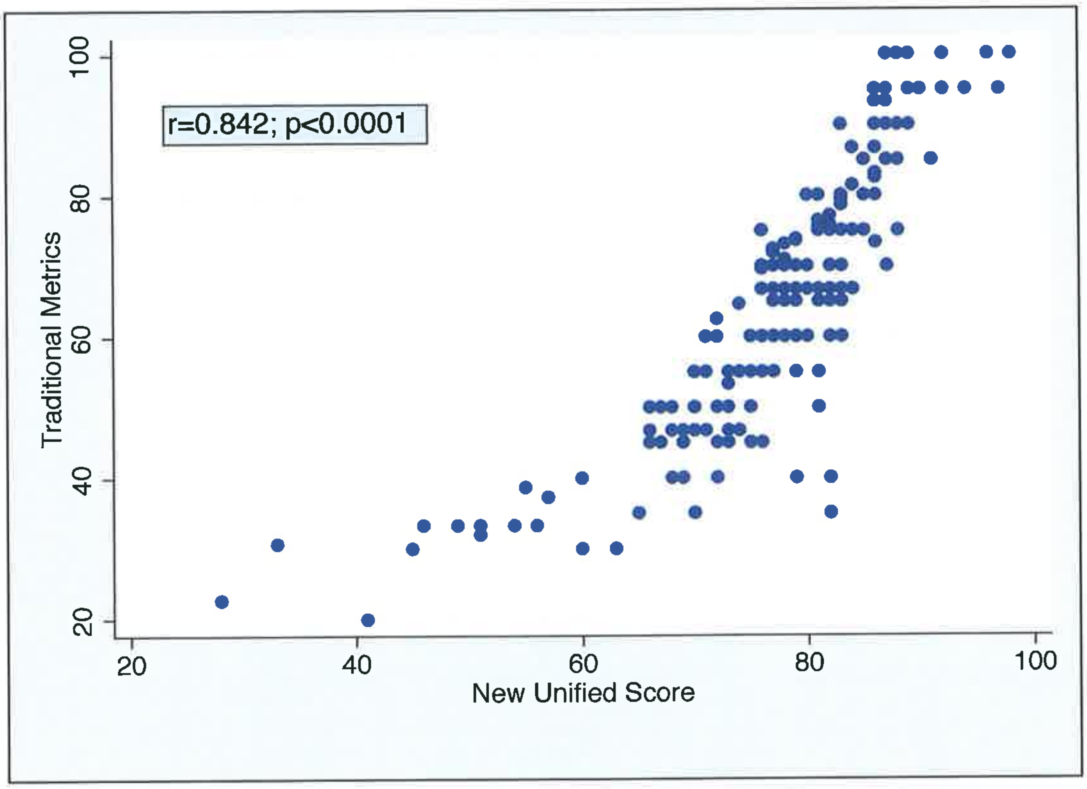


Figure 7.3 Scatterplot displaying correlation between the new unified scores and the objective metrics

7.4.4 Validation of the Formula

The overall co-efficient correlation was $r = 0.82$ ($p<0.0001$) for the path length percentage scores as calculated using the formula compared with the percentile rank scores. Figure 7.4 displays the scatterplot of this correlation and Table 7.4 shows the breakdown of the correlation scores for each individual task. The overall co-efficient correlation was $r = 0.73$ ($p<0.0001$) for the smoothness percentage scores and $r = 0.92$ ($p<0.0001$) for the time percentage scores compared with the percentile rank scores. Figure 7.5 and Figure 7.6 display the scatterplots of the smoothness and time correlations and Table 7.5 and Table 7.6 show the breakdown of the individual score correlations respectively. Percentile rank scores (100 centiles) were calculated from the 214 scores (209 trainees and 5 consultants) for each task for the metrics of time, path length and smoothness.

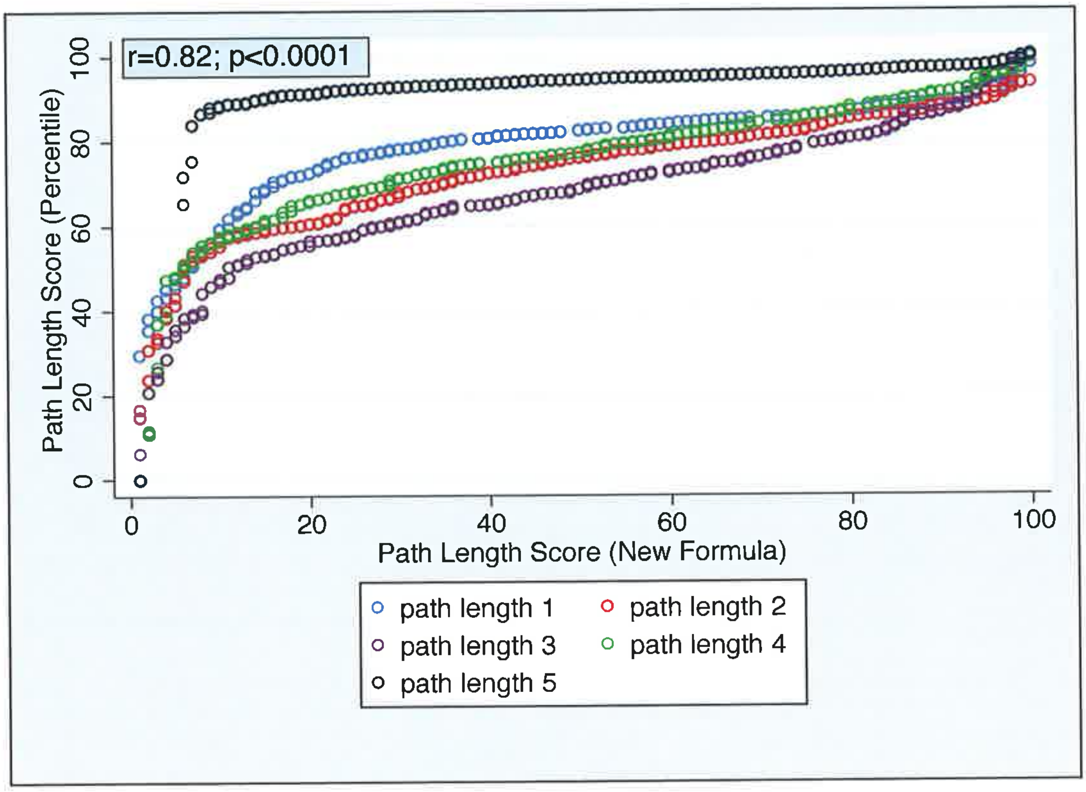


Figure 7.4 Correlation between formula derived path length percentage scores calculated and percentile rank scores

Table 7.4 Correlation between path length percentage scores calculated using the formula and percentile rank scores for each task

N = 214	**Path length percentage scores (formula)**	
Percentile path length scores		
Path Length 1	*r* = 0.8131	p < 0.0001
Path Length 2	*r* = 0.8937	p < 0.0001
Path Length 3	*r* = 0.9238	p < 0.0001
Path Length 4	*r* = 0.8676	p < 0.0001
Path Length 5	*r* = 0.5179	p < 0.0001

Table 7.5 Correlation between smoothness percentage scores calculated using the formula and percentile rank scores for each task

N = 214	Smoothness percentage scores (formula)	
Percentile smoothness scores		
Smoothness 1	$r = 0.7229$	$p < 0.0001$
Smoothness 2	$r = 0.7767$	$p < 0.0001$
Smoothness 3	$r = 0.617$	$p < 0.0001$
Smoothness 4	$r = 0.6625$	$p < 0.0001$
Smoothness 5	$r = 0.7333$	$p < 0.0001$

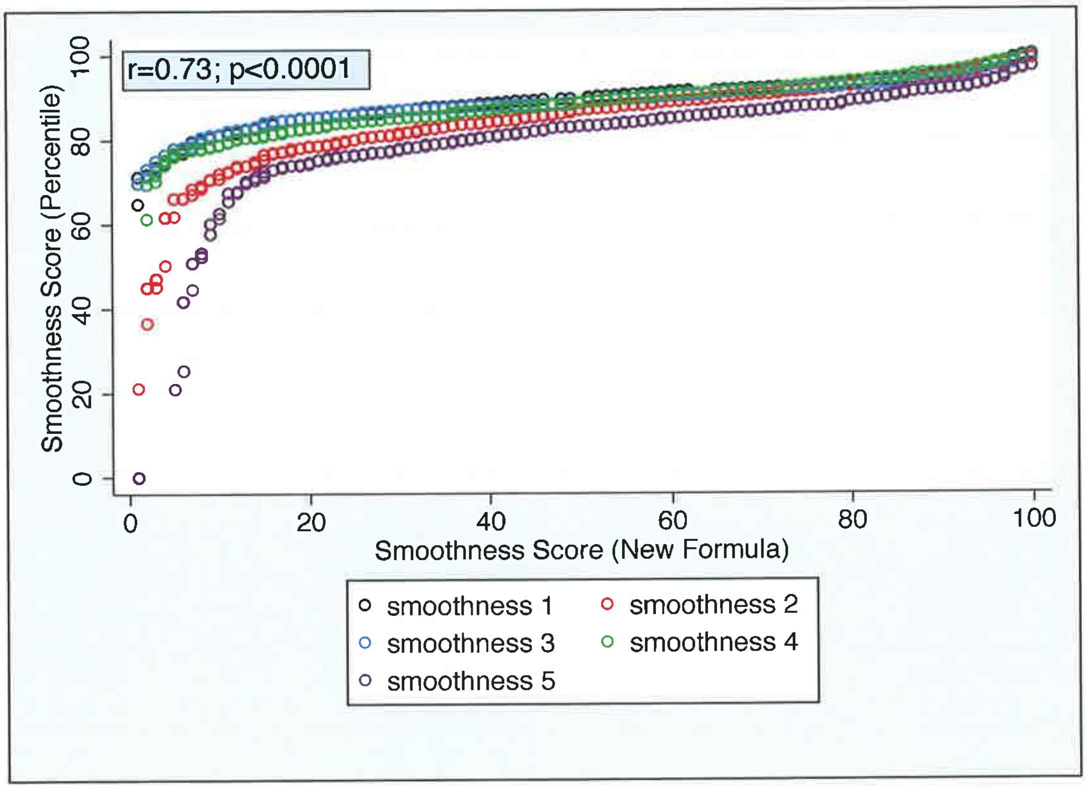


Figure 7.5 Correlation between formula derived smoothness percentage scores and percentile rank scores

Table 7.6 Correlation between time percentage scores calculated using the formula and percentile rank scores for each task

N = 214	Time percentage scores (formula)	
Percentile time scores		
Time 1	$r = 0.9218$	$p < 0.0001$
Time 2	$r = 0.8424$	$p < 0.0001$
Time 3	$r = 0.9238$	$p < 0.0001$
Time 4	$r = 0.9465$	$p < 0.0001$
Time 5	$r = 0.77$	$p < 0.0001$

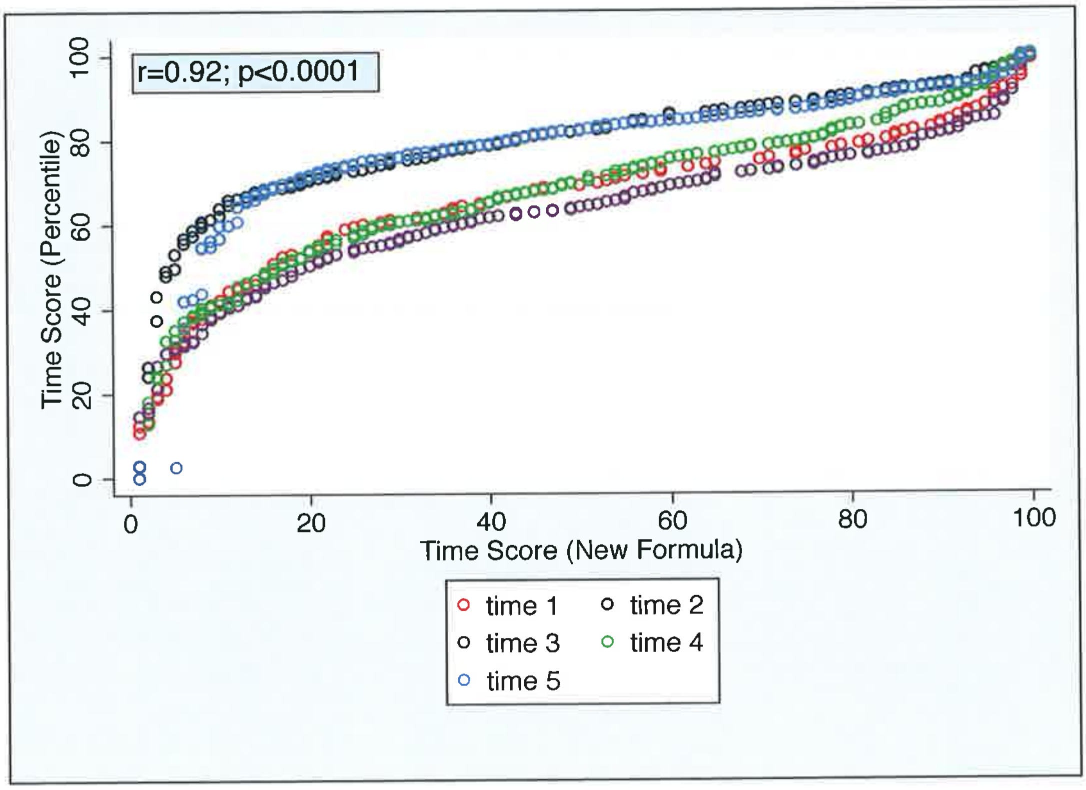


Figure 7.6 Correlation between formula derived time percentage scores and percentile rank scores

7.5 Discussion

Simulators provide us with invaluable objective performance scores for laparoscopic procedures and tasks, however it is challenging to try and apply these scores during formal assessment procedures. This is due to several parameters being recorded and also due to the difference between the units for each metric provided. Also an arbituary path length and smoothness score is useless if it is not compared with an expert score.

The creation of a standarised scoring system for laparoscopic tasks provides trainees with meaningful scores and it allows surgical training bodies to assess trainees during formal skill assessments in a fair and objective manner.

Education systems have different methods of calculating test scores. Simple scoring systems use proportion scores which indicates what proportion of the total marks a person has gained but these scores do not account for factors such as how hard the test is, where a candidate stands in relation to their peers or the margin of error in the test score. The use of standardised scores and percentile ranks are therefore advantageous as test-takers can be compared with a large, nationally representative sample that has taken the test.

By developing a standardised scoring system for the ProMIS simulator we aimed to compare every surgical trainees performance score to a large group of other trainees allowing a fair comparative assessment method. This is more important than ever in the field of surgery. We want to ensure that the process of assessing trainees is fair and transparent and also that those with the highest sores in technical performance are allowed to proceed above those with inferior scores.

In conclusion, our unified score correlates significantly with both expert derived and traditional objective metrics. This scoring method provides a standard of comparing surgical trainees in relation to each other. It also provides user scores, which are easily comprehended and may be used in formal surgical assessments. This standardised scoring system could be applied to a wide range of laparoscopic tasks used in the assessment of surgical trainees.

Chapter 8

General Discussion and Future Work

8.1 Aims of Thesis

The primary aim of this thesis was to evaluate the impact of aptitude (visual spatial, perceptual and psychomotor) upon the rate at which a candidate becomes proficient in laparoscopic procedures. Our secondary aim was to validate novel objective assessment methods of laparoscopic performance.

Summary of Aims:

- To investigate the impact that innate ability has on completion of the learning curve for laparoscopic appendicectomy.
- To investigate the impact that innate ability has on completion of the learning curve for laparoscopic suturing.
- To evaluate the role of self-selection into surgery based on fundamental ability.
- To provide construct and concurrent validity for a novel zone metric used to assess laparoscopic suturing.

- To establish a robust method of comparing laparoscopic performance scores of surgical trainees to expert surgeons by developing a novel objective scoring system for the ProMIS simulator.

8.2 Summary of Main findings

- We evaluated the impact of aptitude on achieving proficiency in laparoscopic appendicectomy in two groups of medical students with grossly different aptitude scores by comparing the number of attempts required to achieve proficiency. The results demonstrated that medical students who have high aptitude achieve proficiency in laparoscopic appendicectomy twice as fast as those with low aptitude.

- When evaluating the impact of aptitude on achieving proficiency in a complex laparoscopic task such as laparoscopic suturing, again those who have high aptitude achieved proficiency much faster than those with low aptitude.

- Not all medical students with low aptitude scores were able to achieve proficiency in laparoscopic procedures. 25% failed to achieve proficiency in laparoscopic appendicectomy and 70% in laparoscopic suturing.

- Aptitude predicted superior baseline performance in objective performance scores as well as a reduced error score for laparoscopic procedures in a group of medical students.

- The visual spatial, psychomotor and perceptual aptitudes of general and plastic surgery HST applicants were compared to a group of medical students interested in pursuing a career in surgery. The results demonstrated that both groups of HST applicants had higher scores in all areas of aptitude than the medical student group.

- Concurrent validity of a novel zone metric was established. There was a significant correlation between the zone metric scores and subjective observer scores. There were significant differences between all three experience groups in the novel zone metric scores, thereby establishing construct validity.

- An objective standardised scoring system was developed for laparoscopic skills on the ProMIS simulator. For all groups (experts, HST's and BST's), there was a significant correlation between the subjective evaluation of the performance and the scores assigned by our standardised system.

8.3 General Conclusions

The principle area that we investigated was the role of aptitude on surgical performance. In the first two experiments, we established that the rate of achieving proficiency for laparoscopic appendicectomy and laparoscopic suturing differs depending on a candidate's aptitude.

Training a surgeon involves considerable time commitments and has significant financial implications. The aim of a surgical training program is to provide a medium to train candidates efficiently, while ensuring that they are proficient in their chosen field at completion of training. As we strive to improve surgical training to ensure minimal attrition rates and achieve technical proficiency upon completion of the program in a time efficient manner, we should select those with the appropriate aptitude. The results of these two experiments confirm that subjects with high aptitude achieve proficiency significantly earlier than those with inferior aptitude.

Not all candidates with low aptitude can achieve proficiency despite multiple attempts. One in four candidates with low aptitude were unable to achieve proficiency in laparoscopic appendectomy despite 18 attempts of the procedure and more than one in two candidates with low aptitude were unable to achieve proficiency in laparoscopic suturing. These findings have significant resource implications.

Work previously carried out by Stefanidis et al (Stefanidis et al., 2006) concluded that the importance of psychomotor testing lies within the prediction of how rapidly one can acquire a laparoscopic skill. Our work specifically aimed at determining the exact difference in rate of skill acquisition between those with contrasting aptitude. We have demonstrated that previously validated aptitude tests can be utilized to predict those who will demonstrate the technical ability and learning facility to achieve proficiency from a ‘novice phase’. Current literature shows that residents with higher visual-spatial scores perform significantly better than did those with lower scores (Wanzel et al., 2002, Wanzel et al., 2003). Our study strongly supports these findings as we have shown that those with high aptitude scores have a reduced error score, superior baseline metric scores and quicker attainment of proficiency.

Our findings also support previous work that demonstrated that not all surgical candidates could achieve proficiency. Grantcharov’s study (Grantcharov and Funch-Jensen, 2009) found that 8.1% of his population group was unable to learn the laparoscopic technique. Our data strongly supports these findings as we also found that there is the proportion of students who fail to progress or follow any upward learning curve.

These experiments were not intended to challenge the belief that “practice makes perfect”, instead it is intended to identify candidates who would require a significant amount of additional training and would struggle in a highly competitive field. We are in a time of dramatic change to the surgical training environment. These findings have significant implications with reference to surgical candidate selection and competency-based progression.

We concluded based on these results that high aptitude predicts a faster learning curve and improved performance in laparoscopic procedures. These findings support the concept of using objective selection processes based on aptitude to select suitable trainees who are likely to flourish in the challenging field of surgery if selected.

It seems appropriate to question how these findings would impact future candidate selection. Before we considered how these findings would be implemented in a real life setting, we felt that a concept worth investigation was the potential self-selection of candidates into the field of surgery. We felt that candidate's were unlikely to continue to pursue a career in surgery if they had self-awareness regarding poor hand eye coordination, manual dexterity and perceptual ability.

From the data analysed in the third experiment (Chapter 5), it is apparent that HST applicants possess fundamental abilities considered important in the field of surgery. It is also clear that they have consistently higher aptitude scores than novice candidates who wish to pursue a career in surgery. This is reassuring and to be encouraged.

Although there appears to be a self-selection process of candidates with high innate ability into surgery, not all of the candidates had high performance scores. There were a small percentage (7.6% for plastic surgery and 11.6% for general surgery) of applicants that underperformed in relation to the medical student population.

This shows us that although the majority of candidates (approximately 90%) naturally possess good fundamental ability appropriate to becoming a competent surgeon, there is a group who did not perform at the expected level in aptitude assessment. This is an important point to consider as this minority group has progressed through basic surgical training, without being identified as potential weak candidates. Therefore it would seem appropriate that the concept of aptitude testing be considered for surgical selection in order to identify this minority group of candidates (approximately 10%).

The secondary focus of this thesis was developing novel objective assessment methods of laparoscopic performance. Proficiency based training has been demonstrated as one of the superior methods of surgical simulation training, but in order to achieve proficiency in a time efficient manner, assessment methods need to be appropriate and robust.

Upon completion of experiment two, one of the criticisms regarding assessment of laparoscopic suturing was that the candidates found it difficult to interpret their scores and how it reflected their performance. It is clear from the results in the laparoscopic suturing experiment that the traditional metrics do assess task progression appropriately, however metrics should provide performance feedback for the user. With this in mind, we developed a novel concept of zone metrics, which calculated the exact point in a 3-D space where one is operating. This space was imagined as 'zones' based on the movements travelled by experienced laparoscopic surgeons while performing a suturing task.

As these new zone metric scores correlated to subjective scores, concurrent validity was established. Construct validity was also established, as there was a statistical difference between three groups with differing laparoscopic experience. By developing this new objective assessment method, it provides us with accurate objective assessment of trainees but also meaningful user scores, which are essential for performance feedback.

Another criticism, which is frequently encountered during assessment of trainees on laparoscopic simulators, is the difficulty in interpreting the metrics produced by the simulator after completing a task or procedure. The ProMIS gives a report of three metrics: time (measured in seconds), path length (measured in millimeters) and smoothness (no units). It is difficult for a trainee to understand their score and to gauge how well they have performed in comparison to an expert score.

We developed a scoring system which combines all three metrics giving a single or unified score between 0 and 100. This was done using a formula which we validated against percentile ranks.

There was a significant correlation with the subjective evaluations of the performances and the scores assigned by our standardised scoring system. This scoring method provides a robust method of comparing novices to experts and provides user scores, which are easily tunderstood and may be used in formal surgical assessments. This standardised scoring system could be applied to a wide range of laparoscopic tasks on any surgical simulator which produces objective metrics.

Overall we believe that these findings have relevance for surgical training. Our findings suggest that aptitude assessment of surgical trainees would be appropriate to identify candidates who would struggle to achieve proficiency targets. It would also be practical to perform these aptitude assessments at an earlier stage of surgical training than at entry to HST. It was demonstrated that 10% of candidates with low aptitude progressed through basic surgical training, which is a significant portion of postgraduate training without being flagged as potential weak candidates. There is also added value for the trainee as it gives them insight into their inherent ability in the context of surgical performance. It also allows them to see what kind of learning pathway would be predicted based on their aptitude and the length of time involved in achieving proficiency in minimally invasive procedures. The next question would then be what weighting should be assigned to aptitude testing in the surgical selection process.

The new metric we have established to assess laparoscopic suturing and the standardised scoring system providing single user metrics which are comparable to expert's performance can be used as a determinant of progression in our surgical training pathway. Surgical training bodies are continuously looking to simulation as a way of bridging the skill gap that trainees are faced with in recent years. In order for simulators to be an effective training tool, feedback must be accurate and valid (Botden et al., 2009a, Boyle et al., 2011). There is also evidence that has demonstrated that providing appropriate feedback can shorten the learning curve for minimally invasive procedures (Xeroulis et al., 2007)

By designing a novel metric that can provide meaningful and objective feedback, we are furthering our surgical curriculum and aiding the learning curve to proficiency. The creation of a standarised scoring system for laparoscopic tasks provides the trainees with meaningful scores and it allows surgical training bodies to objectively assess trainees during formal skill assessments necessary for progression through surgical training.

8.4 Limitations

The findings of chapter 3, 4 and 6 have been demonstrated in a simulated laparoscopic environment. We cannot guarantee that these specific findings would be reproducible in a live setting. However recent evidence seems to support the concept of skills transfer (Gallagher et al., 2013, Buckley, 2013).

Further to this we have stated that those with a higher aptitude have a faster learning curve and can achieve proficiency at a faster rate. The initial part of the learning has been shown to be the period of time when most errors are committed. Although the candidates were penalized if they committed any pre-defined errors, it is not an environment where patient safety is an issue. Therefore, perhaps it is difficult to truly establish if they have overcome the learning curve for a live patient. However it is valid to state that they have achieved proficiency as determined by experts for an index procedure in a simulated environment.

Another limitation is the demonstration of learning curves for those at either end of the aptitude spectrum. The candidates who failed to progress were at the lower end of the aptitude spectrum, and the candidates who excelled were at the highest. However, the majority of surgical candidates are more likely to perform somewhere in between these extremes. By investigating the comparison of aptitude scores of HST applicants and medical students in chapter 5, we aimed to address this limitation.

In chapter 6, it was difficult for the candidates to know whether they were operating in the correct zone, as we did not use overlaying graphics. We felt this would obstruct their view of the operative field. It would be ideal if the screen could flash a different color for example if a candidate went outside the preset zones.

There was a margin of error by setting the in-zone as 0-6 cm and the out-zone as 6-10 cm. If a candidate performed the task 1 cm beyond the appropriate zone then they would be deemed as "not proficient". This was compensated for by assessing for pre-defined errors and by also co-assessing the performances with subjective rating scales. However in order to use zone metric by itself as a method of objective assessment, further investigation and quantification of more accurate zone measurements would be required.

The main limitation of the standardised scoring system was the appropriate weighting of each of the individual metrics. Each laparoscopic task and procedure is different and requires a unique surgical technique. For example for the object positioning task a good time score and path length score is important, however for the sharp dissection task its more important that the candidate takes their time with the task, cuts with precision and has a good smoothness score. We weighted the metrics equally for validation purposes but research into how each task/procedure should be weighted is worth investigating.

A general limitation, which requires discussion, is the concept of validity. Although the concept of validity has been well established in the surgical community, it is constantly evolving. Surgical educators consistently use careful scientific methodology to accomplish this validation goal, however they often base this methodology on an older, previously established framework for showing validity unlike the realm of psychology and educational research. These latter bodies have generally adopted the 1985 consensus standards of the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. These guidelines define validity as "appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores," and validation was defined as the hypothesis driven "process of accumulating evidence to support such inferences." This framework of validity in the standards has changed from "valid instruments" and "types of validity" (construct, face, content, and criterion-related), which was last used in 1974 and absent from their most recent consensus standards published in 1999 (Korndorffer et al., 2010).

A limitation of the thesis is the use of the older framework of validity, which encompasses “valid instruments” and “types of validity” (construct, face, content, and criterion-related). This was used; as it is currently by far the most common way validity is sought in current surgical education research. The reason that the surgical education community has not embraced this new method of establishing validity is unknown. Perhaps like myself, it is done in order to ensure maximal understanding among readers and also to compare with currents standards of surgical educational research. Also it has to be said that several studies are adhering to the concepts of the validity standards without adhering to them in a strict sense. This is an area that will hopefully evolve and gain clarity within the next few years amongst surgical educational literature.

8.5 Future Work

If our results were reproduced in the clinical environment, this would give significant credence to the findings in this thesis.

A specific area that requires further development and validation is establishment of the new zone metric as an assessment of laparoscopic suturing. The programme could be developed to potentially flash a different colour when outside the appropriate zone so that it could truly act as a learning tool that would provide feedback. Further to this, future studies should be carried out to assess if this new metric is a more meaningful way of assessing laparoscopic suturing that traditional metrics. This concept was only hypothesised in this body of work; therefore further testing of this metric would be required in order to give it appropriate substantiation. Another study, which would be worthwhile, is research into the application of the new zone metric for other laparoscopic tasks apart from laparoscopic suturing.

Another specific area of future work, would involve how to weight the metrics for the standardised scoring system. Following on from this would be how to appropriately weight the aptitude scores in the multidimensional assessment used to allow surgical selection and progression.

In this thesis we have explored finding novel ways of surgical assessment using aptitude and the development of new metrics. An area which is currently gaining traction in the field of surgical education and simulation is hand tracking. This is a very exciting concept of which there is currently a paucity of literature. The Tyndall National Research Centre has developed a 'glove' (like a scuba glove), which can be used to track the hand. It is a flat circuit board of sensors built into a glove. By swapping instrument tracking (used in this thesis) for open hand tracking, path analysis of a particular task or procedure could be determined which could help to demonstrate when proficiency is achieved.

Appendix I. OSATS (Observed Structured Assessment of Technical Skill): Reznick *et al.*, 1997

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

Appendix II. Global Rating Scale (Modified OSATS): Grantcharov *et al.*, 2002

Economy of movement					
	1	2	3	4	5
Unnecessary movements	Clear economy of movement and maximum efficiency		Some unnecessary moves		Many unnecessary moves
Confidence of movements	Fluent moves with instruments and no awkwardness		Competent use of instruments but occasionally stiff or awkward		Repeated tentative awkward or inappropriate moves with instruments
Errors					
	1	2	3	4	5
Respect for tissue	Consistently handled tissue appropriately with minimal damage		Handled tissue carefully but occasionally caused inadvertent damage		Frequently used unnecessary orce on tissue or caused damage by inappropriate use of instruments
Precision of operative technique	Fluent, secure and correct technique in all stages of the operative procedure		Careful technique with occasional errors		Imprecise, wrong technique in approaching operative intentions

Appendix III: Visual Spatial Tests

CARD ROTATIONS TEST — S-1 (Rev.)

This is a test of your ability to see differences in figures. Look at the 5 triangle-shaped cards drawn below.

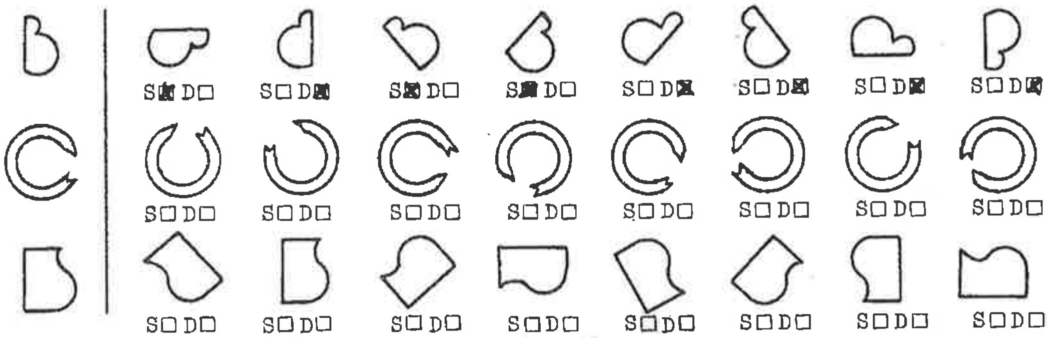
All of these drawings are of the same card, which has been slid around into different positions on the page.

Now look at the 2 cards below:

These two cards are not alike. The first cannot be made to look like the second by sliding it around on the page. It would have to be flipped over or made differently.

Each problem in this test consists of one card on the left of a vertical line and eight cards on the right. You are to decide whether each of the eight cards on the right is the same as or different from the card at the left. Mark the box beside the S if it is the same as the one at the beginning of the row. Mark the box beside the D if it is different from the one at the beginning of the row.

Practice on the following rows. The first row has been correctly marked for you.



Your score on this test will be the number of items answered correctly minus the number answered incorrectly. Therefore, it will not be to your advantage to guess, unless you have some idea whether the card is the same or different. Work as quickly as you can without sacrificing accuracy.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

CUBE COMPARISONS TEST -- S-2 (Rev.)

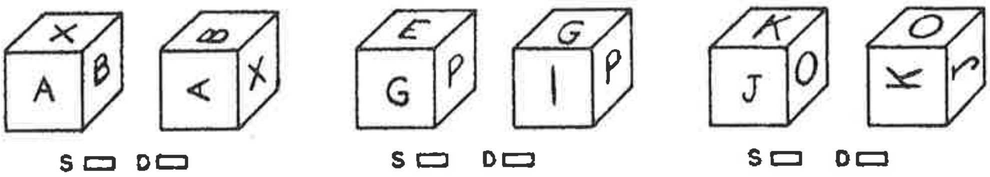
Wooden blocks such as children play with are often cubical with a different letter, number, or symbol on each of the six faces (top, bottom, four sides). Each problem in this test consists of drawings of pairs of cubes or blocks of this kind. Remember, there is a different design, number, or letter on each face of a given cube or block. Compare the two cubes in each pair below.

The first pair is marked D because they must be drawings of <u>different</u> cubes. If the left cube is turned so that the A is upright and facing you, the N would be to the left of the A and hidden, not to the right of the A as is shown on the right hand member of the pair. Thus, the drawings must be of different cubes.

The second pair is marked S because they could be drawings of the same cube. That is, if the A is turned on its side the X becomes hidden, the B is now on top, and the C (which was hidden) now appears. Thus the two drawings could be of the same cube.

<u>Note</u>: No letters, numbers, or symbols appear on more than one face of a given cube. Except for that, <u>any</u> letter, number or symbol can be on the hidden faces of a cube.

Work the three examples below.



The first pair immediately above should be marked D because the X cannot be at the peak of the A on the left hand drawing and at the base of the A on the right hand drawing. The second pair is "different" because P has its side next to G on the left hand cube but its top next to G on the right hand cube. The blocks in the third pair are the same, the J and K are just turned on their side, moving the O to the top.

Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will <u>not</u> be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.

You will have <u>3 minutes</u> for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP.

DO NOT TURN THE PAGE UNTIL YOU ARE ASKED TO DO SO.

MAP PLANNING TEST — SS-3

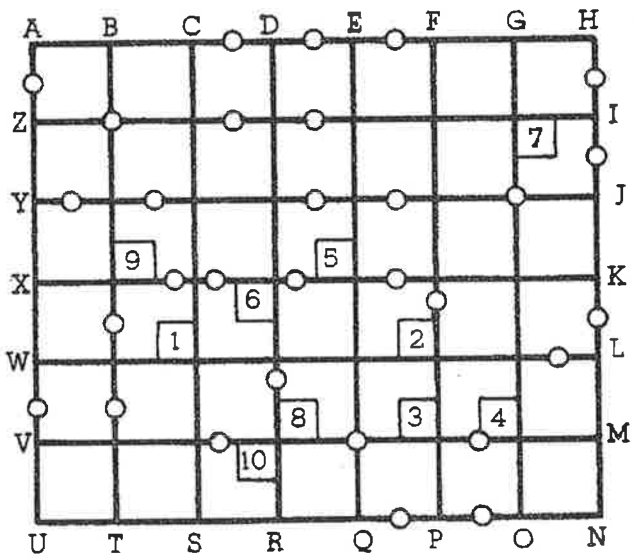
This is a test of your ability to find the shortest route between two places as quickly as possible. The drawing below is a map of a city. The dark lines are streets. The circles are road-blocks, and you cannot pass at the places where there are circles. The numbered squares are buildings. You are to find the shortest route between two lettered points. The number on the building passed is your answer.

Rules: 1. The shortest route will always pass along the side of one and only one of the numbered buildings.

2. A building is not considered as having been passed if a route passes only a corner and not a side.

3. The same numbered building may be used on more than one route.

Look at the sample map below. Practice by finding the shortest route between the various points listed at the right of the map. The first problem has been marked correctly.



The shortest route from:	Passes building:
1. A to Z	1
2. E to S	______
3. P to J	______
4. V to K	______
5. O to F	______
6. G to M	______
7. D to Q	______
8. F to T	______

The answers to the other practice problems are as follows: 2 passes 5; 3 passes 3; 4 passes 2; 5 passes 4; 6 passes 4; 7 passes 6; 8 passes 5.

Your score on this test will be the number of right answers. It will not be to your advantage to guess unless you have some idea which route is correct. Work as rapidly as you can without sacrificing accuracy.

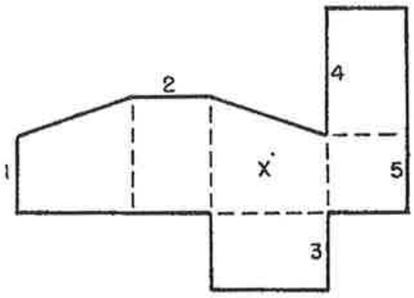
You will have 3 minutes for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

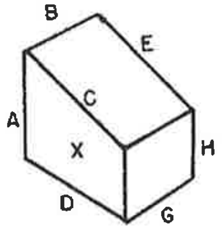
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

SURFACE DEVELOPMENT TEST — VZ-3

In this test you are to try to imagine or visualize how a piece of paper can be folded to form some kind of object. Look at the two drawings below. The drawing on the left is of a piece of paper which can be folded on the dotted lines to form the object drawn at the right. You are to imagine the folding and are to figure out which of the lettered edges on the object are the same as the numbered edges on the piece of paper at the left. Write the letters of the answers in the numbered spaces at the far right.

Now try the practice problem below. Numbers 1 and 4 are already correctly marked for you.





1:	H
2:	
3:	
4:	C
5:	

NOTE: The side of the flat piece marked with the X will always be the same as the side of the object marked with the X. Therefore, the paper must always be folded so that the X will be on the outside of the object.

In the above problem, if the side with edge 1 is folded around to form the back of the object, then edge 1 will be the same as edge H. If the side with edge 5 is folded back, then the side with edge 4 may be folded down so that edge 4 is the same as edge C. The other answers are as follows: 2 is B; 3 is G; and 5 is H. Notice that two of the answers can be the same.

Your score on this test will be the number of correct letters minus a fraction of the number of incorrect letters. Therefore, it will <u>not</u> be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have <u>6 minutes</u> for each of the two parts of this test. Each part has 2 pages. <u>When you have</u> finished Part 1 (pages 2 and 3), STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Appendix IV. FLS (Fundamentals of Laparoscopic Suturing): American College of Surgeons

Parameter	**Score**
Positioning needle in needle holder	1-5
Running needle through suturing pad	1-5
Taking proper bites of the suturing pad while performing the suture	1-5
Throwing thread around the needle holder	1-5
Pulling tight the thread in the proper direction	1-5
Tying the correct surgical knot	1-5

Appendix V. Expert Proficiency Scores

Procedure	Time	Path Length	Smoothness
Laparoscopic Appendicectomy			
Range	(275-557)	(10096-17120)	(1143-2592)
Mean	395	12946	1862
St. Dev.	110.2	2842.6	529.7
Laparoscopic Suturing			
Range	(89-1840)	(3847-5885)	(220-485)
Mean	113	4981	310
St. Dev.	30.7	910.6	84.3

AGGARWAL, R., CROCHET, P., DIAS, A., MISRA, A., ZIPRIN, P. & DARZI, A. 2009. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. *Br J Surg,* 96, 1086-93.

AGGARWAL, R., GRANTCHAROV, T., MOORTHY, K., HANCE, J. & DARZI, A. 2006a. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *Am J Surg,* 191, 128-33.

AGGARWAL, R., GRANTCHAROV, T. P. & DARZI, A. 2007a. Framework for systematic training and assessment of technical skills. *J Am Coll Surg,* 204, 697-705.

AGGARWAL, R., GRANTCHAROV, T. P., ERIKSEN, J. R., BLIRUP, D., KRISTIANSEN, V. B., FUNCH-JENSEN, P. & DARZI, A. 2006b. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann Surg,* 244, 310-4.

AGGARWAL, R., WARD, J., BALASUNDARAM, I., SAINS, P., ATHANASIOU, T. & DARZI, A. 2007b. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg,* 246, 771-9.

AHLBERG, G., ENOCHSSON, L., GALLAGHER, A. G., HEDMAN, L., HOGMAN, C., MCCLUSKY III, D. A., RAMEL, S., SMITH, C. D. & ARVIDSSON, D. 2007. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *American journal of surgery,* 193, 797-804.

ALLEN, J. W., RIVAS, H., COCCHIONE, R. N. & FERZLI, G. S. 2003. Intracorporeal suturing and knot tying broadens the clinical applicability of laparoscopy. *JSLS,* 7**,** 137-40.

AULT, G., REZNICK, R., MACRAE, H., LEADBETTER, W., DAROSA, D., JOEHL, R., PETERS, J. & REGEHR, G. 2001. Exporting a technical skills evaluation technology to other sites. *Am J Surg,* 182**,** 254-6.

BABINEAU, T. J., BECKER, J., GIBBONS, G., SENTOVICH, S., HESS, D., ROBERTSON, S. & STONE, M. 2004. The "cost" of operative training for surgical residents. *Arch Surg,* 139**,** 366-9; discussion 369-70.

BENNETT, C. L., STRYKER, S. J., FERREIRA, M. R., ADAMS, J. & BEART, R. W., JR. 1997. The learning curve for laparoscopic colorectal surgery. Preliminary results from a prospective analysis of 1194 laparoscopic-assisted colectomies. *Arch Surg,* 132**,** 41-4; discussion 45.

BOTDEN, S. M., DE HINGH, I. H. & JAKIMOWICZ, J. J. 2009a. Meaningful assessment method for laparoscopic suturing training in augmented reality. *Surg Endosc,* 23**,** 2221-8.

BOTDEN, S. M., DE HINGH, I. H. & JAKIMOWICZ, J. J. 2009b. Suturing training in Augmented Reality: gaining proficiency in suturing skills faster. *Surg Endosc,* 23**,** 2131-7.

BOYLE, E., AL-AKASH, M., GALLAGHER, A. G., TRAYNOR, O., HILL, A. D. & NEARY, P. C. 2011. Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. *Postgrad Med J,* 87**,** 524-8.

BRIDGES, M. & DIAMOND, D. L. 1999. The financial impact of teaching surgical residents in the operating room. *Am J Surg,* 177, 28-32.

BROE, D., RIDGWAY, P. F., JOHNSON, S., TIERNEY, S. & CONLON, K. C. 2006. Construct validation of a novel hybrid surgical simulator. *Surg Endosc,* 20, 900-4.

BUCKLEY, C. E., KAVANAGH, D. O., TRAYNOR, O. & NEARY, P. C. 2014a. Is the skillset obtained in surgical simulation transferable to the operating theatre? *Am J Surg,* 207, 146-57.

BUCKLEY, C. E., NUGENT, E., NEARY, P. C., TRAYNOR, O. & CARROLL, S. M. 2014b. Do plastic surgical trainees naturally self-select based on fundamental ability? *J Plast Reconstr Aesthet Surg.*

BUCKLEY, C. E. K., D.O.; TRAYNOR, O.; NEARY, P.C. 2013. Is the skillset obtained in surgical simulation transferable to the operating theatre? *Am J Surg,* In Press.

CARROLL, S. M., KENNEDY, A. M., TRAYNOR, O. & GALLAGHER, A. G. 2009. Objective assessment of surgical performance and its impact on a national selection programme of candidates for higher surgical training in plastic surgery. *J Plast Reconstr Aesthet Surg,* 62, 1543-9.

COOPER, C. 1998. Individual differences. *Arnold publishers.*

COWIE, R. 1998. Measurement and modelling of perceived slant in surfaces represented by freely viewed line drawings. *Perception,* 27:505-540.

CROFTS, T. J., GRIFFITHS, J. M., SHARMA, S., WYGRALA, J. & AITKEN, R. J. 1997. Surgical training: an objective assessment of recent changes for a single health board. *BMJ,* 314, 891-5.

CROTHERS, I. 2001. *An experimental analysis of laparoscopic performance with objectively measured psychometric tasks.* . PhD Thesis, Presented to The Queen's University of Belfast.

DELANEY, C. P., CHANG, E., SENAGORE, A. J. & BRODER, M. 2008. Clinical outcomes and resource utilization associated with laparoscopic and open colectomy using a large national database. *Ann Surg,* 247, 819-24.

DELP, S. L., LOAN, J. P., HOY, M. G., ZAJAC, F. E., TOPP, E. L. & ROSEN, J. M. 1990. An interactive graphics-based model of the lower extremity to study orthopaedic surgical procedures. *IEEE Trans Biomed Eng,* 37, 757-67.

DIKMEN, S. S., HEATON, R. K., GRANT, I. & TEMKIN, N. R. 1999. Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *J Int Neuropsychol Soc,* 5, 346-56.

EDENBOROUGH, R. 1994. Using psychometrics; A practical guide to testing and assessment. *Kogan Page Ltd.*

EKSTROM RB, F. J., HARMANN HH. 1976. Manual for kit of factor-referenced cognitive tests. Princeton NJ.

EMAM, T. A., HANNA, G. B., KIMBER, C. & CUSCHIERI, A. 2000. Differences between experts and trainees in the motion pattern of the dominant upper limb during intracorporeal endoscopic knotting. *Dig Surg,* 17, 120-3; discussion 124-5.

ENDE, J. 1983. Feedback in clinical medical education. *JAMA,* 250, 777-81.

ENOCHSSON, L., WESTMAN, B., RITTER, E. M., HEDMAN, L., KJELLIN, A., WREDMARK, T. & FELLANDER-TSAI, L. 2006. Objective assessment of visuospatial and psychomotor ability and flow of residents and senior endoscopists in simulated gastroscopy. *Surg Endosc,* 20, 895-9.

EVERSBUSCH, A. & GRANTCHAROV, T. P. 2004. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. *Surg Endosc,* 18, 1514-8.

FAULKNER, H., REGEHR, G., MARTIN, J. & REZNICK, R. 1996. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med,* 71, 1363-5.

FIGERT, P. L., PARK, A. E., WITZKE, D. B. & SCHWARTZ, R. W. 2001. Transfer of training in acquiring laparoscopic skills. *J Am Coll Surg,* 193, 533-7.

FRIED, G. M., FELDMAN, L. S., VASSILIOU, M. C., FRASER, S. A., STANBRIDGE, D., GHITULESCU, G. & ANDREW, C. G. 2004. Proving the value of simulation in laparoscopic surgery. *Ann Surg,* 240, 518-25; discussion 525-8.

GALLAGHER, A. G., COWIE, R., CROTHERS, I., JORDAN-BLACK, J. A. & SATAVA, R. M. 2003. PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc,* 17, 1468-71.

GALLAGHER, A. G., LEONARD, G. & TRAYNOR, O. J. 2009. Role and feasibility of psychomotor and dexterity testing in selection for surgical training. *ANZ J Surg,* 79, 108-13.

GALLAGHER, A. G., MCCLURE, N., MCGUIGAN, J., RITCHIE, K. & SHEEHY, N. P. 1998. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy,* 30, 617-20.

GALLAGHER, A. G., RITTER, E. M., CHAMPION, H., HIGGINS, G., FRIED, M. P., MOSES, G., SMITH, C. D. & SATAVA, R. M. 2005. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg,* 241, 364-72.

GALLAGHER, A. G. & SATAVA, R. M. 2002. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. *Surg Endosc,* 16, 1746-52.

GALLAGHER, A. G., SEYMOUR, N. E., JORDAN-BLACK, J. A., BUNTING, B. P., MCGLADE, K. & SATAVA, R. M. 2013. Prospective, randomized assessment of transfer of training (ToT) and transfer effectiveness ratio (TER) of virtual reality simulation training for laparoscopic skill acquisition. *Ann Surg,* 257, 1025-31.

GILLIAM, A. D. 2009. Construct validity of the ProMIS laparoscopic simulator. *Surg Endosc,* 23, 1150.

GRANTCHAROV, T. P., BARDRAM, L., FUNCH-JENSEN, P. & ROSENBERG, J. 2003. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *Am J Surg,* 185, 146-9.

GRANTCHAROV, T. P. & FUNCH-JENSEN, P. 2009. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. *Am J Surg,* 197, 447-9.

GRANTCHAROV, T. P., KRISTIANSEN, V. B., BENDIX, J., BARDRAM, L., ROSENBERG, J. & FUNCH-JENSEN, P. 2004. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg,* 91, 146-50.

GRANTCHAROV, T. P. & REZNICK, R. K. 2008. Teaching procedural skills. *BMJ,* 336, 1129-31.

GUMBS, A. A., HOGLE, N. J. & FOWLER, D. L. 2007. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. *J Am Coll Surg,* 204, 308-13.

GURUSAMY, K., JUNNARKAR, S., FAROUK, M. & DAVIDSON, B. R. 2008. Meta-analysis of randomized controlled trials on the safety and effectiveness of day-case laparoscopic cholecystectomy. *Br J Surg,* 95, 161-8.

GURUSAMY, K. S., AGGARWAL, R., PALANIVELU, L. & DAVIDSON, B. R. 2009. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database of Systematic Reviews.*

HAMILTON, E. C., SCOTT, D. J., KAPOOR, A., NWARIAKU, F., BERGEN, P. C., REGE, R. V., TESFAY, S. T. & JONES, D. B. 2001. Improving operative performance using a laparoscopic hernia simulator. *American journal of surgery,* 182, 725-728.

HASSAN, I., GERDES, B., KOLLER, M., DICK, B., HELLWIG, D., ROTHMUND, M. & ZIELKE, A. 2007. Spatial perception predicts laparoscopic skills on virtual reality laparoscopy simulator. *Childs Nerv Syst,* 23, 685-9.

HEDMAN, L., STROM, P., ANDERSSON, P., KJELLIN, A., WREDMARK, T. & FELLANDER-TSAI, L. 2006. High-level visual-spatial ability for novices correlates with performance in a visual-spatial complex surgical simulator task. *Surg Endosc,* 20, 1275-80.

I., K. 2001. Learning from Bristol: the report of the public enquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995.

KEHLET, H. 2008. Fast-track colorectal surgery. *Lancet,* 371, 791-3.

KERR, B. & O'LEARY, J. P. 1999. The training of the surgeon: Dr. Halsted's greatest legacy. *Am Surg,* 65, 1101-2.

KOHN LT, C. J., DONALDSON M 1999. To Err is HUman: BUilding a Safer Health System. Washington DC: Institude of Medicine.

KOLOZSVARI, N. O., FELDMAN, L. S., VASSILIOU, M. C., DEMYTTENAERE, S. & HOOVER, M. L. 2011a. Sim one, do one, teach one: considerations in designing training curricula for surgical simulation. *J Surg Educ,* 68, 421-7.

KOLOZSVARI, N. O., KANEVA, P., BRACE, C., CHARTRAND, G., VAILLANCOURT, M., CAO, J., BANASZEK, D., DEMYTTENAERE, S., VASSILIOU, M. C., FRIED, G. M. & FELDMAN, L. S. 2011b. Mastery versus the standard proficiency target for basic laparoscopic skill training: effect on skill transfer and retention. *Surg Endosc,* 25, 2063-70.

KORNDORFFER, J. R., DUNNE, J. B., SIERRA, R., STEFANIDIS, D., TOUCHARD, C. L. & SCOTT, D. J. 2005. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *Journal of the American College of Surgeons,* 201, 23-29.

KORNDORFFER, J. R., JR., KASTEN, S. J. & DOWNING, S. M. 2010. A call for the utilization of consensus standards in the surgical education literature. *Am J Surg,* 199, 99-104.

LARSEN, C. R., SOERENSEN, J. L., GRANTCHAROV, T. P., DALSGAARD, T., SCHOUENBORG, L., OTTOSEN, C., SCHROEDER, T. V. & OTTESEN, B. S. 2009. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *BMJ,* 338, b1802.

LIN, Y. Y., SHABBIR, A. & SO, J. B. 2010. Laparoscopic appendectomy by residents: evaluating outcomes and learning curve. *Surg Endosc,* 24, 125-30.

LL, K. 1970. *As told to Robert B. Parke. The Pilot Maker.,* New York, Grosset and Dunlap.

LUCAS, S. M., ZELTSER, I. S., BENSALAH, K., TUNCEL, A., JENKINS, A., PEARLE, M. S. & CADEDDUT, J. A. 2008. Training on a Virtual Reality

Laparoscopic Simulator Improves Performance of an Unfamiliar Live Laparoscopic Procedure. *Journal of Urology,* 180, 2588-2591.

MARTIN, J. A., REGEHR, G., REZNICK, R., MACRAE, H., MURNAGHAN, J., HUTCHISON, C. & BROWN, M. 1997. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg,* 84, 273-8.

MASON, J. D., ANSELL, J., WARREN, N. & TORKINGTON, J. 2013. Is motion analysis a valid tool for assessing laparoscopic skill? *Surg Endosc,* 27, 1468-77.

MCCLUSKY, D. A., GALLAGHER, A. G., RITTER, E. M., LEDERMAN, A. B., VAN SICKLE, K. R., BAGHAI, M. & SMITH, C. D. 2004. Virtual reality training improves junior residents' operating room performance: Results of a prospective, randomized, double-blinded study of the complete laparoscopic cholecystectomy. *Journal of the American College of Surgeons,* 199, S73-S73.

METALIS, S. 1985. Effects of massed versus distributed practice on acquisition of video game skill. *Percept Motor Skills* 457–458.

MOORE, M. J. & BENNETT, C. L. 1995. The learning curve for laparoscopic cholecystectomy. The Southern Surgeons Club. *Am J Surg,* 170, 55-9.

MOULTON, C. A., DUBROWSKI, A., MACRAE, H., GRAHAM, B., GROBER, E. & REZNICK, R. 2006. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. *Ann Surg,* 244, 400-9.

NEARY, P. C., BOYLE, E., DELANEY, C. P., SENAGORE, A. J., KEANE, F. B. & GALLAGHER, A. G. 2008. Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; results from the first course for experienced senior laparoscopic surgeons. *Surg Endosc,* 22, 2301-9.

NUGENT, E. 2011. *The evaluation of fundamental ability in acquiring minimally invasive surgical skill sets.* MD Thesis, Presented to The Royal College of Surgeons in Ireland.

NUGENT, E., HSEINO, H., BOYLE, E., MEHIGAN, B., RYAN, K., TRAYNOR, O. & NEARY, P. 2012a. Assessment of the role of aptitude in the acquisition of advanced laparoscopic surgical skill sets : Results from a virtual reality-based laparoscopic colectomy training programme. *Int J Colorectal Dis,* 27, 1207-14.

NUGENT, E., HSEINO, H., BOYLE, E., MEHIGAN, B., RYAN, K., TRAYNOR, O. & NEARY, P. 2012b. Assessment of the role of aptitude in the acquisition of advanced laparoscopic surgical skill sets: results from a virtual reality-based laparoscopic colectomy training programme. *Int J Colorectal Dis,* 27, 1207-14.

NUGENT, E., JOYCE, C., PEREZ-ABADIA, G., FRANK, J., SAUERBIER, M., NEARY, P., GALLAGHER, A. G., TRAYNOR, O. & CARROLL, S. 2012c. Factors influencing microsurgical skill acquisition during a dedicated training course. *Microsurgery,* 32, 649-56.

NUGENT, E., SHIRILLA, N., HAFEEZ, A., O'RIORDAIN, D. S., TRAYNOR, O., HARRISON, A. M. & NEARY, P. 2013. Development and evaluation of a simulator-based laparoscopic training program for surgical novices. *Surg Endosc,* 27, 214-21.

PALTER, V. N., ORZECH, N., REZNICK, R. K. & GRANTCHAROV, T. P. 2013. Validation of a structured training and assessment curriculum for technical skill acquisition in minimally invasive surgery: a randomized controlled trial. *Ann Surg,* 257, 224-30.

PERKINS, N., STARKES, J. L., LEE, T. D. & HUTCHISON, C. 2002. Learning to use minimal access surgical instruments and 2-dimensional remote visual feedback: how difficult is the task for novices? *Adv Health Sci Educ Theory Pract,* 7, 117-31.

REZNICK, R., REGEHR, G., MACRAE, H., MARTIN, J. & MCCULLOCH, W. 1997. Testing technical skill via an innovative "bench station" examination. *Am J Surg,* 173, 226-30.

RITTER, E. M., KINDELAN, T. W., MICHAEL, C., PIMENTEL, E. A. & BOWYER, M. W. 2007. Concurrent validity of augmented reality metrics applied to the fundamentals of laparoscopic surgery (FLS). *Surg Endosc,* 21, 1441-5.

ROGERS, D. A., ELSTEIN, A. S. & BORDAGE, G. 2001. Improving continuing medical education for surgical techniques: applying the lessons learned in the first decade of minimal access surgery. *Ann Surg,* 233, 159-66.

ROSENTHAL, M. E., RITTER, E. M., GOOVA, M. T., CASTELLVI, A. O., TESFAY, S. T., PIMENTEL, E. A., HARTZLER, R. & SCOTT, D. J. 2010. Proficiency-based Fundamentals of Laparoscopic Surgery skills training results in durable performance improvement and a uniform certification pass rate. *Surg Endosc,* 24**,** 2453-7.

ROSSER, J. C., ROSSER, L. E. & SAVALGI, R. S. 1997. Skill acquisition and assessment for laparoscopic surgery. *Arch Surg,* 132**,** 200-4.

SARKER, S. K., CHANG, A. & VINCENT, C. 2006. Technical and technological skills assessment in laparoscopic surgery. *JSLS,* 10**,** 284-92.

SATAVA, R. M. 1993. Virtual reality surgical simulator. The first steps. *Surg Endosc,* 7**,** 203-5.

SATAVA, R. M. 1999. Emerging technologies for surgery in the 21st century. *Arch Surg,* 134**,** 1197-202.

SAUERLAND, S., JASCHINSKI, T. & NEUGEBAUER, E. A. 2010. Laparoscopic versus open surgery for suspected appendicitis. *Cochrane Database Syst Rev,* CD001546.

SCOTT, D. J., BERGEN, P. C., REGE, R. V., LAYCOCK, R., TESFAY, S. T., VALENTINE, R. J., EUHUS, D. M., JEYARAJAH, D. R., THOMPSON, W. M. & JONES, D. B. 2000. Laparoscopic training on bench models: Better and more cost effective than operating room experience? *Journal of the American College of Surgeons,* 191**,** 272-283.

SEYMOUR, N. E., GALLAGHER, A. G., ROMAN, S. A., O'BRIEN, M. K., BANSAL, V. K., ANDERSEN, D. K. & SATAVA, R. M. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg,* 236**,** 458-63; discussion 463-4.

SIMONS, A. J., ANTHONE, G. J., ORTEGA, A. E., FRANKLIN, M., FLESHMAN, J., GEIS, W. P. & BEART, R. W., JR. 1995. Laparoscopic-assisted colectomy learning curve. *Dis Colon Rectum,* 38, 600-3.

SKIDMORE, F. D. 1997. Junior surgeons are becoming deskilled as result of Calman proposals. *BMJ,* 314, 1281.

SMITH, C. D., FARRELL, T. M., MCNATT, S. S. & METREVELI, R. E. 2001. Assessing laparoscopic manipulative skills. *Am J Surg,* 181, 547-50.

SNYDER, C. W., VANDROMME, M. J., TYRA, S. L., PORTERFIELD, J. R., JR., CLEMENTS, R. H. & HAWN, M. T. 2011. Effects of virtual reality simulator training method and observational learning on surgical performance. *World J Surg,* 35, 245-52.

STEFANIDIS, D., ACKER, C. & HENIFORD, B. T. 2008. Proficiency-based laparoscopic simulator training leads to improved operating room skill that is resistant to decay. *Surg Innov,* 15, 69-73.

STEFANIDIS, D., HOPE, W. W., KORNDORFFER, J. R., JR., MARKLEY, S. & SCOTT, D. J. 2010. Initial laparoscopic basic skills training shortens the learning curve of laparoscopic suturing and is cost-effective. *J Am Coll Surg,* 210, 436-40.

STEFANIDIS, D., KORNDORFFER, J. R., JR., BLACK, F. W., DUNNE, J. B., SIERRA, R., TOUCHARD, C. L., RICE, D. A., MARKERT, R. J., KASTL, P. R. & SCOTT, D. J. 2006. Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery,* 140, 252-62.

STYLOPOULOS, N., COTIN, S., DAWSON, S., OTTENSMEYER, M., NEUMANN, P., BARDSLEY, R., RUSSELL, M., JACKSON, P. & RATTNER, D. 2003. CELTS: a clinically-based Computer Enhanced Laparoscopic Training System. *Stud Health Technol Inform,* 94, 336-42.

STYLOPOULOS, N., COTIN, S., MAITHEL, S. K., OTTENSMEYE, M., JACKSON, P. G., BARDSLEY, R. S., NEUMANN, P. F., RATTNER, D. W. & DAWSON, S. L. 2004. Computer-enhanced laparoscopic training system (CELTS): bridging the gap. *Surg Endosc,* 18, 782-9.

SWEET, R. M., HANANEL, D. & LAWRENZ, F. 2010. A unified approach to validation, reliability, and education study design for surgical technical skills training. *Arch Surg,* 145, 197-201.

SZABO, Z., HUNTER, J., BERCI, G., SACKIER, J. & CUSCHIERI, A. 1994. Analysis of surgical movements during suturing in laparoscopy. *Endosc Surg Allied Technol,* 2, 55-61.

TEKKIS, P. P., SENAGORE, A. J., DELANEY, C. P. & FAZIO, V. W. 2005. Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections. *Ann Surg,* 242, 83-91.

TRAXER, O., GETTMAN, M. T., NAPPER, C. A., SCOTT, D. J., JONES, D. B., ROEHRBORN, C. G., PEARLE, M. S. & CADEDDU, J. A. 2001. The impact of intense laparoscopic skills training on the operative performance of urology residents. *J Urol,* 166, 1658-61.

VAILLANCOURT, M., GHADERI, I., KANEVA, P., VASSILIOU, M., KOLOZSVARI, N., GEORGE, I., SUTTON, F. E., SEAGULL, F. J., PARK, A. E., FRIED, G. M. & FELDMAN, L. S. 2011. GOALS-incisional hernia: a valid assessment of simulated laparoscopic incisional hernia repair. *Surg Innov,* 18, 48-54.

VAN BRUWAENE, S., DE WIN, G. & MISEREZ, M. 2009. How much do we need experts during laparoscopic suturing training? *Surg Endosc,* 23, 2755-61.

VAN DONGEN, K. W., AHLBERG, G., BONAVINA, L., CARTER, F. J., GRANTCHAROV, T. P., HYLTANDER, A., SCHIJVEN, M. P., STEFANI, A., VAN DER ZEE, D. C. & BROEDERS, I. A. 2011a. European consensus on a competency-based virtual reality training program for basic endoscopic surgical psychomotor skills. *Surg Endosc,* 25, 166-71.

VAN DONGEN, K. W., AHLBERG, G., BONAVINA, L., CARTER, F. J., GRANTCHAROV, T. P., HYLTANDER, A., SCHIJVEN, M. P., STEFANI, A., VAN DER ZEE, D. C. & BROEDERS, I. A. M. J. 2011b. European consensus on a competency-based virtual reality training program for basic endoscopic surgical psychomotor skills. *Surgical Endoscopy and Other Interventional Techniques,* 25, 166-171.

VAN HERZEELE, I., O'DONOGHUE, K. G., AGGARWAL, R., VERMASSEN, F., DARZI, A. & CHESHIRE, N. J. 2010. Visuospatial and psychomotor aptitude predicts endovascular performance of inexperienced individuals on a virtual reality simulator. *J Vasc Surg,* 51, 1035-42.

VAN HOVE, P. D., TUIJTHOF, G. J., VERDAASDONK, E. G., STASSEN, L. P. & DANKELMAN, J. 2010. Objective assessment of technical surgical skills. *Br J Surg,* 97, 972-87.

VASSILIOU, M. C., FELDMAN, L. S., ANDREW, C. G., BERGMAN, S., LEFFONDRE, K., STANBRIDGE, D. & FRIED, G. M. 2005. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg,* 190, 107-13.

VERDAASDONK, E. G., DANKELMAN, J., LANGE, J. F. & STASSEN, L. P. 2008. Transfer validity of laparoscopic knot-tying training on a VR simulator to a realistic environment: a randomized controlled trial. *Surg Endosc,* 22, 1636-42.

VERDAASDONK, E. G., STASSEN, L. P., VAN WIJK, R. P. & DANKELMAN, J. 2007. The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. *Surg Endosc,* 21, 214-9.

WALKER M, P. J. 1998. Teaching in Theatre. *Teaching and Learning in Medical Practice.* Rickmansworth.

WANZEL, K. R., HAMSTRA, S. J., ANASTAKIS, D. J., MATSUMOTO, E. D. & CUSIMANO, M. D. 2002. Effect of visual-spatial ability on learning of spatially-complex surgical skills. *Lancet,* 359, 230-1.

WANZEL, K. R., HAMSTRA, S. J., CAMINITI, M. F., ANASTAKIS, D. J., GROBER, E. D. & REZNICK, R. K. 2003. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery,* 134, 750-7.

WATSON, D. I., BAIGRIE, R. J. & JAMIESON, G. G. 1996. A learning curve for laparoscopic fundoplication. Definable, avoidable, or a waste of time? *Ann Surg,* 224, 198-203.

WHERRY, D. C., MAROHN, M. R., MALANOSKI, M. P., HETZ, S. P. & RICH, N. M. 1996. An external audit of laparoscopic cholecystectomy in the steady state performed in medical treatment facilities of the Department of Defense. *Ann Surg,* 224, 145-54.

WIGGINS G, M. J. 2001. *Understanding by Design,* Upper Saddle River, NJ, Prentice Hall, Inc.

WOLFE, B. M., SZABO, Z., MORAN, M. E., CHAN, P. & HUNTER, J. G. 1993. Training for minimally invasive surgery. Need for surgical skills. *Surg Endosc,* 7, 93-5.

XEROULIS, G. J., PARK, J., MOULTON, C. A., REZNICK, R. K., LEBLANC, V. & DUBROWSKI, A. 2007. Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery,* 141, 442-9.