Human-centered design in laparoscopic skills acquisition
Shifting paradigms in the age of technology

Ganni, S.

DOI
10.4233/uuid:b597524f-3321-465e-90da-9963bd9f7754

Publication date
2020

Document Version
Final published version

Citation (APA)
Ganni, S. (2020). *Human-centered design in laparoscopic skills acquisition: Shifting paradigms in the age of technology*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:b597524f-3321-465e-90da-9963bd9f7754

HUMAN CENTRED DESIGN IN LAPAROSCOPIC SKILLS ACQUISITION

SHIFTING PARADIGMS IN THE AGE OF TECHNOLOGY

SANDEEP GANNI

Human-centred Design in Laparoscopic Skills Acquisition

Shifting Paradigms in the Age of Technology

Dissertation

for the purpose of obtaining the degree of
doctor at Delft University of Technology,
by the authority of the Rector Magnificus,
Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board of the Doctorates
to be defended publicly on
Thursday 27 August 2020 at 15:00

by

Sandeep Ganni
Master of Business Administration in Healthcare Management, Coventry University
born in Rajahmundry, India

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairman
Prof. dr. ir. R. H. M. Goossens	Delft University of Technology, promotor
Prof. dr. J. J. Jakimowicz	Delft University of Technology, promotor
Dr. M. K. Chmarra	Delft University of Technology, copromotor
Dr. S. M. B. I. Botden	Radboud University Medical Centre
Independent Members:	
Prof. dr. N. Bouvy	AMC Maastricht
Prof. dr. D. R. Pai	Mahatma Gandhi Medical College, India
Prof. dr. T. J. M. van der Cammen	Delft University of Technology
Prof. ir. D. J. van Eijk	Delft University of Technology

This research was funded by GSL Medical College, Rajahmundry, India under the GSL Smart Lab project for institutional implementation of simulation based curricula.

Printed by: Print.Amsterdam BV

Cover design: Sandeep Ganni

ISBN 978-16-362-5776-1

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Tell me and I'll forget;
Show me and I may remember;
Involve me and I will understand.

Benjamin Franklin

CONTENTS

Summary

Minimal Access Surgery (MAS) has multi-faceted implications on the different stake holders involved in implementation. It requires the surgeon to cope with the ergonomic and cognitive challenges required to perform a surgical procedure. It needs the surgical team to work coherently for the smooth functioning of the technologically complex operating room (OR). It needs educators and policy makers to embrace reform into the novel practices of training skills in MAS. It needs the industry to engage in research to bring out newer tools and technologies akin to training tools. These stakeholders are interdependent and the culmination of their efforts effecting the outcomes of quality of patient care and ultimately contributing to enhanced patient safety and reduced morbidity and mortality.

The successful implementation of MAS and the envisioned outcomes of patient safety are brought about by training surgeons and surgical teams using comprehensive training curricula encompassing technical skills and non-technical skills using appropriate assessment protocols. MAS is a valuable technique, provided that a number of requirements are met. These range from basic ergonomic knowledge to complex serious adverse events. Worldwide, it is estimated that 10 percent of surgical cases are readmitted due to injuries resulting from medical errors and over 80 percent of surgeons are suffering from some form of lower back and neck injuries due to strained ergonomic postures resulting either from faulty posture or prolonged operating times. These complications can be attributed to several factors, the most important being the lack of adequate training in MAS according to a recent survey.

This thesis explores the premise of implementation of MAS in the areas of curriculum design, implementation of training protocols, assessment tools and emerging technologies and trends that improve the learning curve and the transfer of skills from a skills lab setting to the OR. It is essential for surgeons to focus on therapeutic skills and patient safety, whilst not being distracted due to insufficient knowledge of MAS specific instruments and equipment and the accompanying technology. Coping with these MAS specific tools requires a structured curriculum. This curriculum needs multi-modal elements in order to reduce the unfamiliarity with complex instrument interfaces and translates the true therapeutic proficiency of a

surgeon into procedural performance. **Chapter 1** addresses the complexity of skills required for MAS and enumerates these under technical, cognitive and other non-technical skills. Further, the tools available for the surgeons to train these skills are discussed under several modalities discussing the benefit of one over the other and the need for an inclusive protocol that combines the best of every modality relevant for training. The assessment protocols are addressed under the different assessment modes: expert, self, peer and VR or augmented simulator. The benefits of each of these modes are discussed.

Based on the literature study, **Chapter 2** aims to determine the current standards and protocols followed around the world in MAS training and accreditation. A survey was conducted that received 663 responses from 73 countries resulting in a global visualization of different types of training courses, the quality of academic and training materials used and the standards for accreditation. A varied and segregated approach to implementation of training was observed with resulting in limited implementation of curricula. This was supported by the response that most participants who attended such courses expressed interest in accredited courses and furthering their MAS skills. The general consensus in the survey revealed an assorted interpretation of the concept of curriculum.

Chapter 3 outlines the role of self-assessment in MAS training curricula as a measure of not only the proficiency gained by the trainee but also the effectiveness of the curriculum in delivering the desired outcomes. This was measured by the interrelation of expert and self-assessment. The study found self-assessment to be on par with expert assessment in most areas of training. However, there were some inconsistencies in correlation between the expert assessment and the self-assessment. This was reason to reevaluate how the curriculum was designed and how self-assessment is used. This reevaluation is done in **Chapter 4**. Deciphering and deconstructing self-assessment in terms of practical and professional implications prior to the assessment led to greater correlation in scores between experts and trainees. Furthermore, the overall performance of trainees was found to be relatively higher when these modified training protocols in self-assessment were applied, demonstrating trainees' holistic understanding of evaluation criteria and reaching intended goals. **Chapter 5** explores the use of a modified competency assessment tool for laparoscopic suturing exercises, the LS-CAT for objective evaluation of suturing skills in a training setting. The resulting tool was proposed as an equivalent to the objective structured assessment of technical skills (OSATS) which has been used with success in clinical practice but has not been demonstrated to provide formative assessment of individual technical skills crucial in training curricula. The LS-CAT was designed by deconstructing the component tasks of advanced suturing technique into four subcategories to enable task specific scoring and feedback. It demonstrated excellent inter-rater reliability and identification of expertise level along the progression of learning curves during the training curriculum.

Chapter 6 explains that, while assessment in training curricula is an important facet of skills acquisition, translating these assessment skills into clinical practice is another facet. The use of advanced virtual reality (VR) systems and augmented reality (AR) systems has enabled and conditioned trainees and assessors to objectively and consistently evaluate surgical performance using advanced metrics and tools. However, translation of these objective evaluation criteria into clinical assessment is limited by use of additional equipment and lack of retrospective evaluation. Chapter 6 furthermore explores the use of a software-based motion tracking tool for objective evaluation of surgical performance in an operating theatre. Common metrics among high fidelity VR simulators and those of the software were compared and analyzed to assess performance. The resulting study depicted the software's ability to clearly distinguish between expert and novice performance. However, the metrics should only be considered as one aspect of performance as they do not truly reflect the procedural proficiency of the surgeon or the task but provide an objective measure in the movement of instruments in a manner similar to experts. This required proving the validity and usability of software in clinical practice. The study summarized in **Chapter 7** aims to validate the software in a blinded trial comprising of experts and novices and further propose thresholds for proficiency in such metrics using motion tracking. The software demonstrated validity in the blinded trial and an algorithm was proposed which was a ratio of path length, average distance and jerk index. This algorithm proved to distinctively differentiate an expert from a novice performance paving the way for psychomotor skills assessment in the OR. This was proposed as an inclusive tool for psychomotor skills assessment in the developing trend of artificial intelligence-based image and error recognition.

The success of any training curricula is measured by the short- and long-term effects it has on the participants. The short-term quality indices that determine a curriculum's effectiveness are its ability to progress learning curve and its ability to make all trainees ready for clinical performance in the procedure of training. Similarly, a curriculums long-term effects are determined by the post-curriculum benefits; especially clinical performance, retention and development of skills obtained. **Chapter 8** outlines a three-year multi-variate study conducted in India to observe the benefits of a multi-modal curriculum; the Laparoscopic Surgical Skills (LSS) curriculum followed by a study group in comparison to traditional apprenticeship model of training followed by a control group. The metrics used for evaluation of the curriculum's effectiveness traversed not only the surgical performance of trainees but also considered patient outcomes, operating times, post-operative complications and average cost per patient. The results showed significant differences in surgical skills acquisition amongst the study and control group. The study group demonstrated a faster acquisition and progression of learning curve required for MAS and outperformed the control group in areas of patient and hospital outcomes. Another facet to successful curricula is the adoption of newer training methodologies and tools in tone with changing trends in educational theories and strategies. This requires a survey of gap-analysis and trainee needs. **Chapter 9** analyses participant surveys from and before the onset of this study spanning from years 2013 to 2018. 173 responses were received during this period with participants rating the curriculum consistently better with the changes implemented in the curriculum. The areas where the participants scored better were in correlation to the changes implemented in the curriculum in self-assessment and training material pertaining to it.

While technical skills acquisition is the objective of many VR and AR simulators currently in use, the advent of head mounted VR goggles have made it easier to employ immersive environments for recreating OR scenarios and incorporating surgical teams for training in non-technical skills. Current simulators used in skills labs lack the real-life disruptions in surgical processes and distractions that occur due to team interactions and ambient noises in the OR. One argument to this notion is surgeons need to be trained and assessed in environments similar to that of real-life situations for the training and assessment to be effective. This ensures effective transfer of skills from the skills lab to clinical practice and potentially increases the trainee's adaption to the OR environment. The study in **Chapter 10** aims to evaluate the disruptions to surgical flow in the OR in different complexity of surgical procedures and the impact it has on a surgeon's mental resources in terms of stress, attention and fatigue. The aim of this study being incorporating these findings into head mounted VR simulators to create immersive environments that engage surgeons in skills labs. Objective data was gathered using skin sensors that measured galvanic skin response, heat flux and metabolic task equivalent during various surgical process disruptions and distractions. Further in **Chapter 11** the validity and usability of head mounted VR simulators were studied in a group of expert and novice surgeons in India. Several standard measures were used that included presence questionnaire, heuristics, questionnaire for intuitive use and the NASA task load index. Increased mental load at a cognitive level was observed in novices in comparison to experts inferring that expert surgeons are equipped to deal with disruptions and novice surgeons are benefitted from the training before operating in the OR.

Finally, **Chapter 12** summarizes the interrelation and parallels between all the research in this thesis in the direction of human centered design in laparoscopic skills acquisitions and further explores on the projects that have stemmed from this research and the future prospects of training in MAS.

Samenvatting

Minimaal invasieve chirurgie (een MAS-ingreep) heeft voor de verschillende belanghebbenden die bij de implementatie betrokken zijn nogal wat implicaties. Het vereist dat de chirurg de ergonomische en de cognitieve uitdagingen, die nodig zijn om een chirurgische procedure uit te kunnen voeren, aangaat. Het chirurgische team moet voor een soepele werking van de technologische complexe operatiekamer (OK) coherent werken. Het heeft opleiders en beleidsmakers nodig om de hervormingen in de nieuwe praktijken van de opleidingsvaardigheden in MAS te omarmen. Het heeft de industrie nodig om onderzoek te kunnen doen om nieuwere hulpmiddelen en technologieën, die lijken op de hulpmiddelen tijdens de training, te kunnen introduceren. Deze belanghebbenden zijn onderling afhankelijk en het resultaat van hun inspanningen zullen de resultaten van de kwaliteit van de patiëntenzorg beïnvloeden en uiteindelijk bijdragen tot een grotere patiëntveiligheid en verminderde morbiditeit en mortaliteit.

Een succesvolle implementatie van MAS en de beoogde resultaten van patiëntveiligheid worden bewerkstelligd door het opleiden van chirurgen en chirurgische teams met behulp van geschikte beoordelingsprotocollen en door gebruik te maken van uitgebreide trainingsprogramma's, die technische en niet-technische vaardigheden omvatten. MAS is, mits aan een aantal eisen wordt voldaan, een waardevolle techniek. Deze eisen variëren van elementaire ergonomische kennis tot complexe ernstige bijwerkingen. Geschat wordt dat er wereldwijd 10% van de chirurgische patiënten als gevolg van medische fouten en verwondingen opnieuw moet worden opgenomen en meer dan 80% van de chirurgen als gevolg van een verkeerde houding of langdurig opereren aan een vorm van onderrug- of nekletsel lijdt. Deze complicaties kunnen worden toegeschreven aan verschillende factoren, waarvan volgens een recent onderzoek de belangrijkste het gebrek aan een adequate opleiding in minimaal invasieve chirurgie is.

Dit proefschrift onderzoekt het uitgangspunt van de implementatie van minimaal invasieve chirurgie op het gebied van curriculumontwerp, de implementatie van trainingsprotocollen, de beoordelingsinstrumenten en de opkomende technologieën en trends, die de leercurve en de overdracht van vaardigheden van een skillab omgeving naar de OK verbeteren. Het is voor chirurgen essentieel om zich, zonder afgeleid te worden door onvoldoende kennis van MAS-specifieke instrumenten en apparatuur en de bijbehorende technologie, te kunnen concentreren op de therapeutische vaardigheden en de patiëntveiligheid. Kunnen omgaan met deze MAS-specifieke hulpmiddelen vereist een gestructureerd curriculum. Dit curriculum heeft om de onbekendheid met complexe instrument interfaces te verminderen multimodale elementen nodig en vertaalt de echte therapeutische vaardigheid van een chirurg in procedurele prestaties.

Hoofdstuk 1 behandelt de complexiteit van de vaardigheden die voor MAS zijn vereist en somt deze onder technische, cognitieve en andere niet-technische vaardigheden op. Verder worden de hulpmiddelen die de chirurgen beschikbaar hebben om deze vaardigheden te trainen onder verschillende modaliteiten besproken, waarbij het voordeel van de één boven de ander wordt besproken en de behoefte aan een inclusief protocol, dat het beste van elke modaliteit combineert met wat relevant is voor de training. De beoordelingsprotocollen komen in de verschillende beoordelingsmodi aan bod: expert, self, peer en VR of augmented simulator. De voordelen van elk van deze modi zullen worden besproken.

Op basis van literatuuronderzoek beoogt **Hoofdstuk 2** de huidige normen en protocollen te bepalen, die wereldwijd bij de MAS-training en accreditatie worden gevolgd. Er werd een enquête gehouden waarop 663 reacties uit 73 landen kwam, wat resulteerde in een wereldwijde visualisatie van verschillende soorten trainingen, de kwaliteit van het gebruikte academische en trainingsmateriaal en de normen voor accreditatie. Er werd een gevarieerde en gesegregeerde benadering van de implementatie van training waargenomen, wat resulteerde in een beperkte implementatie van curricula. Dit werd ondersteund door de reactie dat de meeste deelnemers aan dergelijke cursussen belangstelling toonden voor geaccrediteerde cursussen en het verbeteren van hun MAS-vaardigheden. De algemene consensus in de enquête bracht een diverse interpretatie van het concept leerplan aan het licht.

Hoofdstuk 3 schetst de rol van de zelfevaluatie in de leerplannen voor MAS-opleidingen als maatstaf voor niet alleen de bekwaamheid die de student heeft verworven, maar ook voor de effectiviteit van het leerplan bij het behalen van de gewenste resultaten. Dit werd gemeten door de onderlinge relatie van deskundige en zelfevaluatie. Uit het onderzoek bleek dat de zelfbeoordeling op de meeste opleidingsgebieden op hetzelfde niveau als de beoordeling door een deskundige lag. Er waren echter in de correlatie tussen de beoordeling door deskundigen en de zelfbeoordeling enkele inconsistenties. Dit was reden om opnieuw te evalueren hoe het curriculum is ontworpen en hoe de zelfevaluatie wordt gebruikt.

Deze herevaluatie wordt in **Hoofdstuk 4** gedaan. Het ontcijferen en het deconstrueren van de zelfbeoordeling in termen van praktische en professionele implicaties voorafgaand aan de beoordeling leidde tot een grotere correlatie tussen de scores van de experts en de stagiaires. Bovendien bleken de algehele prestaties van de stagiaires relatief hoger te zijn wanneer deze aangepaste trainingsprotocollen bij de zelfevaluatie werden toegepast, wat het holistische begrip van de stagiaires van de evaluatiecriteria en het bereiken van de beoogde doelen aantoont.

Hoofdstuk 5 onderzoekt het gebruik van een aangepast competentie beoordelingsinstrument voor laparoscopische hechtingsoefeningen, de LS-CAT voor objectieve evaluatie van hechtvaardigheden in een trainingsomgeving. Het resulterende hulpmiddel werd voorgesteld als een equivalent van de objectieve gestructureerde beoordeling van technische vaardigheden (OSATS) die met succes in de klinische praktijk is gebruikt, maar waarvan niet is aangetoond dat het een formatieve beoordeling biedt van de individuele technische vaardigheden, die in leerplannen cruciaal zijn. De LS-CAT is ontworpen door de componenttaken van een geavanceerde hechttechniek in vier subcategorieën te deconstrueren om taakspecifieke scores en feedback mogelijk te maken. Het toonde een uitstekende interbeoordelaarsbetrouwbaarheid en een identificatie van het expertiseniveau langs de voortgang van leercurven tijdens het trainings curriculum.

Hoofdstuk 6 legt uit dat hoewel beoordeling in leerplannen een belangrijk facet van het verwerven van vaardigheden is, het vertalen van deze beoordelingsvaardigheden naar de klinische praktijk een ander facet is. Het gebruik van geavanceerde virtual reality-systemen (VR) en augmented reality-systemen (AR) heeft stagiaires en assessoren in staat gesteld en geconditioneerd om chirurgische prestaties met behulp van geavanceerde metrische gegevens en hulpmiddelen objectief en consistent te kunnen evalueren. De vertaling van deze objectieve evaluatiecriteria in een klinische beoordeling wordt echter beperkt door het gebruik van aanvullende apparatuur en het ontbreken van een evaluatie achteraf. Hoofdstuk 6 verkent verder het gebruik van een softwaregebaseerd hulpmiddel voor bewegingsregistratie voor een objectieve evaluatie van chirurgische prestaties in een operatiekamer. Gemeenschappelijke statistieken onder hifi VR-simulatoren en die van de software werden vergeleken en geanalyseerd om de prestaties te kunnen beoordelen. Het resulterende onderzoek illustreerde het vermogen van de software om duidelijk onderscheid te kunnen maken tussen de prestaties van deskundigen en beginners. De metrische gegevens mogen echter slechts als één aspect van de prestatie worden beschouwd, aangezien ze niet echt de procedurele bekwaamheid van de chirurg of de taak weerspiegelen, maar een objectieve maatstaf vormen voor de beweging van de instrumenten op een manier die vergelijkbaar is met die van de deskundigen. Dit vereiste het aantonen van de validiteit en de bruikbaarheid van software in de klinische praktijk.

Het onderzoek dat in **Hoofdstuk 7** is samengevat, heeft tot doel om de software in een geblindeerde proef met deskundigen en beginners te valideren en verder drempelwaarden voor te stellen voor bekwaamheid in dergelijke statistieken met behulp van bewegingsregistratie. De software toonde in de geblindeerde proef validiteit aan en er werd een algoritme voorgesteld dat een verhouding tussen de padlengte, de gemiddelde afstand en de ruk index was. Dit algoritme bleek een deskundige van een beginner te kunnen onderscheiden en wat de weg vrijmaakte voor de beoordeling van psychomotorische vaardigheden in de OK. Dit werd voorgesteld als een alomvattend hulpmiddel voor de beoordeling van psychomotorische vaardigheden in de zich ontwikkelende trend van op kunstmatige intelligentie gebaseerde beeld- en foutherkenning.

Het succes van een trainingsprogramma wordt afgemeten aan de korte- en de langetermijneffecten die het op de deelnemers heeft. De kwaliteitsindexen op korte termijn, die de effectiviteit van een curriculum bepalen, zijn het vermogen om de leercurve te verbeteren en het vermogen om alle cursisten klaar te stomen voor klinische prestaties in de trainingsprocedure. Evenzo worden de langetermijneffecten van een curriculum bepaald door de voordelen na het curriculum; vooral de klinische prestaties en het behoud en de ontwikkeling van de verworven vaardigheden.

Hoofdstuk 8 schetst een driejarig multi-variate onderzoek, dat in India werd uitgevoerd om de voordelen van een multimodaal curriculum te observeren: tijdens het curriculum laparoscopische chirurgische vaardigheden (LSS) werd een studiegroep in vergelijking met het traditionele leermodel van training gevolgd door een controlegroep. De metrische gegevens, die werden gebruikt voor de evaluatie van de effectiviteit van het curriculum, hadden niet alleen betrekking op de chirurgische prestaties van de cursisten, maar hielden ook rekening met de resultaten van de patiënt, de operatietijden, de postoperatieve complicaties en de gemiddelde kosten per patiënt. De resultaten lieten significante verschillen in de verwerving van de chirurgische vaardigheden tussen de studie- en controlegroep zien. De studiegroep vertoonde een snellere acquisitie en progressie van de leercurve, die nodig is voor MAS en presteerde op het gebied van de patiënt- en ziekenhuisresultaten beter dan de controlegroep. Een ander facet van de succesvolle curricula is de toepassing van nieuwere trainingsmethodologieën en -instrumenten, in overeenstemming met de veranderende trends in de onderwijstheorieën en -strategieën. Dit vereist een overzicht van de gap-analyse en de behoeften van de stagiares.

Hoofdstuk 9 analyseert de deelnemersonderzoeken van en voor het begin van dit onderzoek dat zich uitstrekt van 2013 tot 2018. In deze periode werden 173 reacties ontvangen, waarbij deelnemers het curriculum met de veranderingen die in het curriculum werden doorgevoerd consistent beter beoordeelden. De gebieden waarop de deelnemers beter scoorden, waren gecorreleerd met de veranderingen, die in het curriculum voor de zelfevaluatie en het trainingsmateriaal dat erop betrekking had, werden geïmplementeerd.

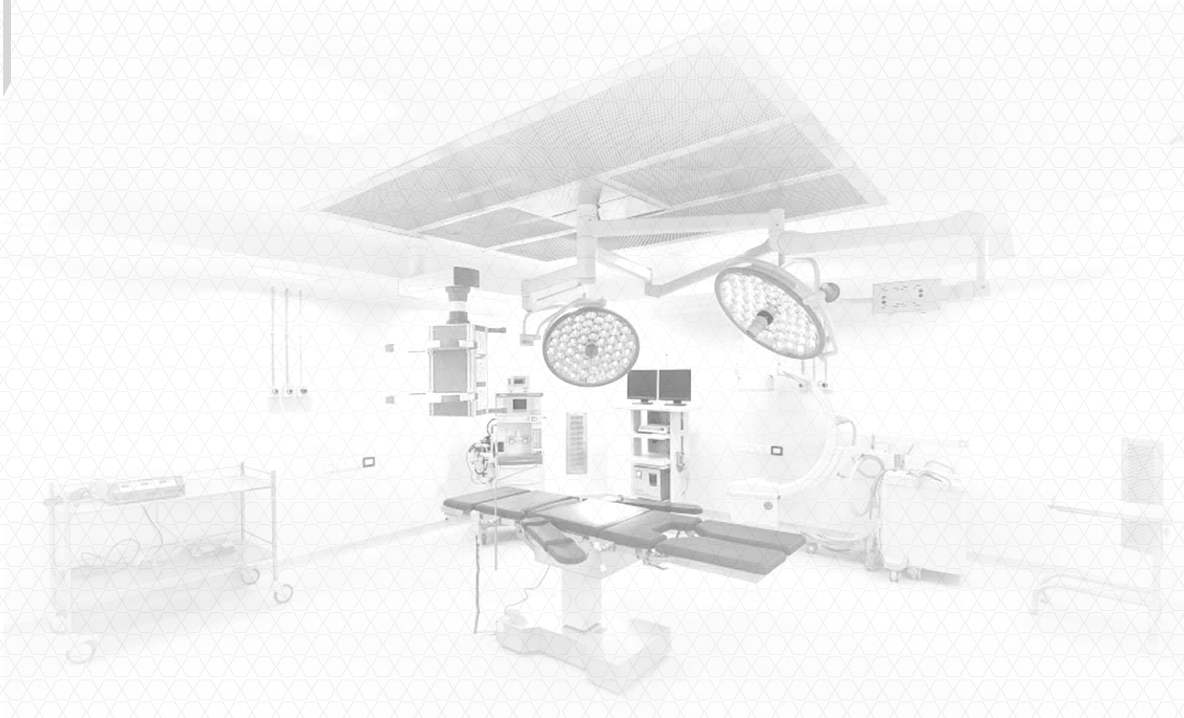
Hoewel het verwerven van technische vaardigheden het doel van veel VR- en AR-simulatoren is, die momenteel in gebruik zijn, heeft de komst van een op het hoofd gemonteerde VR-bril het gemakkelijker gemaakt om immersieve omgevingen voor het recreëren van OK-scenario's en het opnemen van chirurgische teams voor training in niet-technische vaardigheden te gebruiken. De huidige simulatoren, die worden gebruikt in skillab omgevingen, missen de echte verstoringen in de chirurgische processen en de afleidingen die optreden als gevolg van teaminteracties en omgevingsgeluiden in de OK. Een argument voor dit idee is om de training en de beoordeling effectief te laten zijn, moeten chirurgen getraind en beoordeeld worden in omgevingen, die vergelijkbaar zijn met die in de praktijk. Dit zorgt voor een effectieve overdracht van de vaardigheden van de skillab omgeving naar de klinische praktijk en het verhoogt mogelijk ook de aanpassing van de stagiair aan de omgeving van de OK.

Het onderzoek in **Hoofdstuk 10** heeft tot doel om de verstoringen van de chirurgische flow in de OK te beoordelen in verschillende soorten van complexiteit van chirurgische procedures en de impact die het heeft op de mentale middelen van een chirurg in termen van stress, aandacht en vermoeidheid. Het doel van dit onderzoek is om deze bevindingen in de op het hoofd gemonteerde VR-simulatoren te integreren om meeslepende omgevingen te creëren en chirurgen in skillab omgevingen te betrekken. Objectieve gegevens werden verzameld met behulp van huidsensoren, die tijdens verschillende verstoringen en afleidingen van het chirurgische proces de galvanische huidreactie, de warmteflux en de metabole taakequivalent meten.

Verder werden in **Hoofdstuk 11** de validiteit en de bruikbaarheid van de VR-simulatoren op het hoofd van een groep van deskundige en beginnende chirurgen in India bestudeerd. Er werden verschillende standaardmetingen gebruikt, waaronder aanwezigheidsvragenlijsten, heuristieken, vragenlijsten voor intuïtief gebruik en de taakbelastingindex van de NASA. Verhoogde mentale belasting op cognitief niveau werd in vergelijking met de deskundigen bij beginners waargenomen, in de veronderstelling dat deskundige chirurgen zijn uitgerust om met verstoringen om te gaan en dat beginnende chirurgen baat hebben bij de training voordat ze in de OK gaan opereren.

Hoofdstuk 12 vat tenslotte de onderlinge samenhang en de parallellen samen tussen al het onderzoek in dit proefschrift in de richting van mensgericht ontwerpen bij het verwerven van laparoscopische vaardigheden en gaat dieper op de projecten in, die voortkomen uit dit onderzoek en de toekomstperspectieven voor de opleiding in minimaal invasieve chirurgie.

1



INTRODUCTION AND OUTLINE

1

INTRODUCTION

Minimal Access Surgery (MAS) has revolutionised the field of general surgery by means of a radical shift in procedural, equipment and technique requirements justified by the benefits it offers to patient safety and recovery.[1–5] The training required to perform MAS has also significantly departed from the traditional Halstedian apprenticeship model used for laparotomy.[6, 7] Surgeons are required to cope with two-dimensional representation of the operative field, hand-eye coordination, fulcrum effect of fixed instruments and limited range of movement.[8–13] In addition, ever changing technology, patient safety and ethical concerns inspired educators to perform MAS training outside the operating room.[6] The challenges that plague MAS in its current state are ergonomic problems, operating room inefficiencies, surgical variability and workforce challenges that affect the outcome of value-based healthcare.[14–18] MAS skills are known to widely vary among surgeons and are often associated with incidence of post-operative complications, re-operation rates, re-admissions or visits to the emergency department.[16, 19] As it stands, a wide range of training modalities and equipment is available for surgeons to train. Unfortunately, training is mostly performed outside structured curricula.[20] This thesis aims to evaluate current training practices and explore developing tools for optimal training in MAS.

1

NEED FOR TRAINING IN MAS

MAS calls for a unique set of technical and non-technical surgical skills different from that of laparotomy. Heemskerk et al [21] enumerates technical skills surgeon has to cope with as:

1. Two-dimensional (2D) vision using a conventional monitor reduces perception of depth.
2. A disturbed eye–hand–target axis decreases ergonomics and dexterity.
3. The long, rigid instruments used in laparoscopic surgery magnify the surgeon's natural hand tremor.
4. The rigid instruments with four (out of six) degrees of freedom limit the surgeon's natural range of motion, decreasing dexterity.
5. Fixed abdominal entry points result in limited freedom of motion and movement of the tip of the instrument to the opposite direction of the outer part of the instrument, a technical drawback known as the fulcrum effect.
6. Camera instability increases fatigue.
7. Limited tactile feedback decreases dexterity.

Whereas non-technical skills can be broadly classified based on the NOTECHS classification [22] as:

1. Leadership and management
2. Teamwork and cooperation
3. Problem solving and decision making
4. Situation awareness

Acquiring these skills has been proven to improve learning curves.[23] Surveys of surgical residency directors and surgical trainees strongly indicate (<90%) that there is an imminent need for MAS training outside the operating room (OR).[20, 24] Moreover, due to the nature of skills involved in MAS, skills can no longer be evaluated under the auspices of good hands or subjective evaluation. Objective evaluation is now the standard for skills assessment.[25–27] The impact of training skills on patient morbidity and mortality, financial implications of errors and training residents cannot be

Figure 1.1: Animal model setup in wet lab

Figure 1.2: Video box trainer with adjustable base and monitor

overlooked in value-based healthcare delivery. For this reason, training technical and non-technical skills in MAS has been explored extensively outside the OR to improve surgical performance and patient safety.

TRAINING TECHNICAL SKILLS

Technical skills can be trained in MAS in two categories namely task or procedural and scenario-based training. In task training, component tasks and skills are deciphered and addressed, whereas in scenario-based training, complete procedures are replicated. There are several tools available to simulate both task and scenario-based training. They can be broadly categorised into live animals, human cadavers, animal/cadaver models, video-based box trainers, computer based virtual reality simulators. (Figure 1.3)

Human cadavers and animal models have been long used for both laparotomy and open surgery. In MAS training however, additional setup is required. The abdominal wall is either insufflated or parts of the tissue are harvested and incorporated into box trainers. (Figure 1.1) Despite the fact these models are the cornerstone for surgical training due to their high fidelity; ethical concerns, costs and availability to a wider audience have made them less than ideal for mainstream training.[28] Moreover, the performance assessments done on these models are almost always subjective in nature.

Box trainers are designed to mimic abdominal wall and are fitted with a camera or an attachment for an endoscopic camera and a monitor. (Figure 1.2) Port openings are provided for laparoscopic instruments insertion typical to those of procedure specific requirements and mimic the fulcrum effect. Often these box trainers are placed on

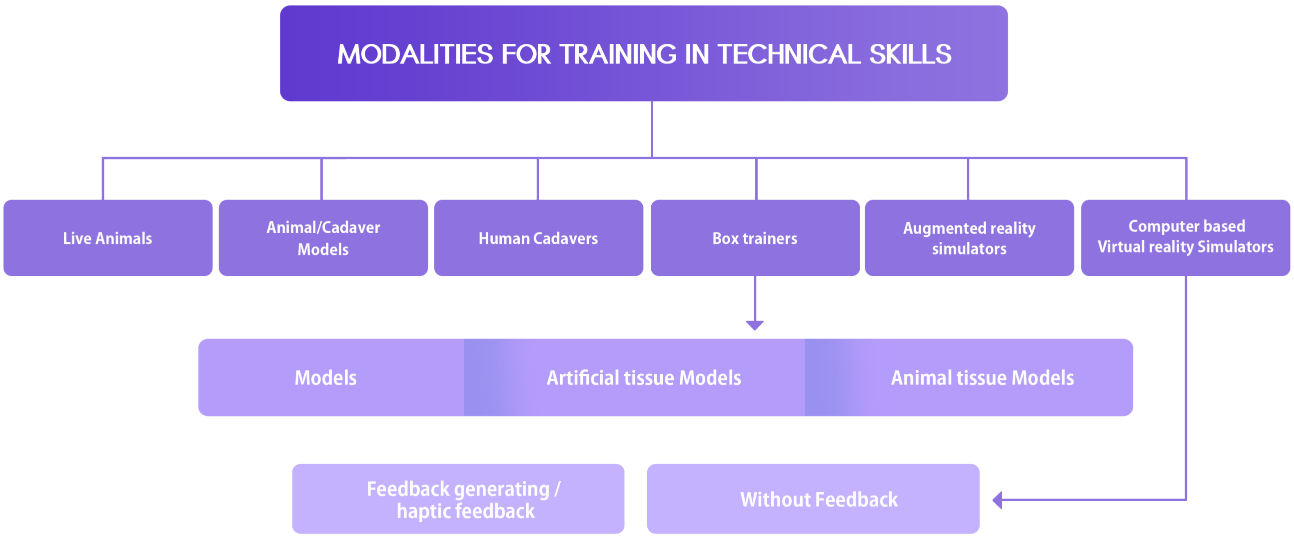


Figure 1.3: Classification of training modalities in MAS

1

(a) The TrEndo system

(b) The ProMIS Simulator

Figure 1.4: Simulator systems

height adjustable tables to further recreate true ergonomics. These box trainers aid in developing basic psycho-motor skills to advanced suturing techniques subject to the quality of materials and models incorporated in them.[29] Similar to animal and cadaver models traditional box trainers are limited to subjective assessment.

Augmented reality (AR) simulators on the other hand overlay computer generated images onto live video feed. The use of real instruments and tissue or objects generate true haptic feedback. They build on the benefits of the box trainers and have additional sensors and tracking equipment to generate objective feedback that is crucial in MAS training.[30] Several products are available that offer different degrees of tracking. TrEndo developed by the Delft University offers an affordable solution that can be retrofitted to box trainers by using gimbals and optical sensors to track instrument movement.[31] (Figure 1.4a) The ProMIS augmented reality simulators are tailored to operate in a custom-made enclosure with inbuilt sensors and trackers.[32] (Figure 1.4b) Newer AR simulators are incorporating semi VR overlay on the physical models to make a hybrid simulator that replicates true haptics whilst being visually realistic to tissue.[33] (Figure 1.5)

Computer based virtual reality (VR) simulators have gained a lot of attention due to their novelty and advanced objective assessment and training protocols. VR simulators offer flexibility in various aspects of training. Users have the opportunity to vary the

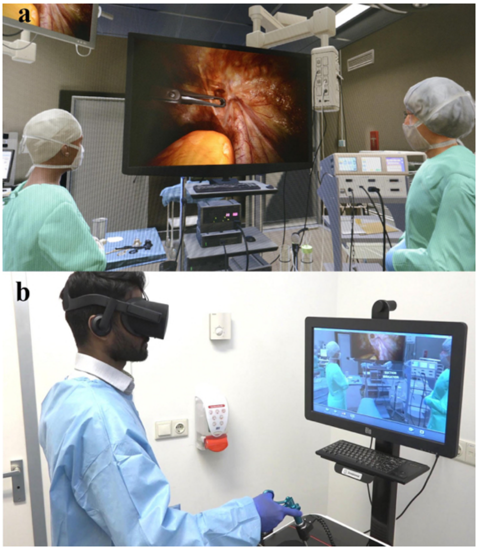
Figure 1.5: The Laparo advanced training station with VR overlay

difficulty from novice to advanced tasks, change pathology of the tissue from normal to atypical anatomy and enable repetitive practice.[34] Furthermore, VR simulators are equipped with an extensive array of sensors and tracking devices that make objective assessment of surgical skills feasible. Novice surgeons can train on these simulators without the need for expert observers and can track their progress over time. The built-in software offers detailed instructions on usage and procedural techniques and tasks guided by interactive and intuitive feedback. One of the significant drawbacks of VR simulators was the lack of haptic feedback and thus the lack of realism. Haptic feedback or force feedback has been proven to be one of the most essential aspects for VR simulators to be effective in terms of both trainee engagement and validity of task assessment and precision therein.[35] However, most modern simulators in the market offer haptic feedback by using tensiometers and actuators.[36] Several VR simulators are available like the LapMentor by 3D systems, LapSim by SurgicalScience, Lap-X by Medical-X. (Figure 1.6a)

Never versions of these simulators now offer true VR experience employing head mounted displays. These work in sync with existing interface of the simulators and provide a 360-degree environment of the OR setup and personnel inside. (Figure 1.6b)

Despite the numerous choices available for training technical skills in MAS each of the modes of training hold their value over one another. For example, suturing exercises performed in a box trainer are far more effective in training over the VR simulators.[37]

(a) The LapMentor VR simulator by Simbionix



(b) The Virtual OR setup with head mounted display

Figure 1.6: Virtual Reality systems

Whereas learning suturing exercises in a VR simulator is easier since it enables self-learning and step-by-step guided visual instructions on the interface despite the lack of proper haptic feedback.[38] Similarly animal and animal cadaver models though considered the holy grail of surgical training cannot offer the exact pathological representation of the anatomy required to perform a procedure as opposed to a VR simulator which can easily replicate anatomy. Human cadavers offer perfect representation of anatomy and in some cases even pathology. The VR simulator, however, allows infinite repetitive procedural training and parameters allowing objective assessment.

TRAINING NON-TECHNICAL SURGICAL SKILLS

Non-technical surgical skills as detailed above, are a combination of cognitive and interpersonal skills. Whilst most attention in surgical skills is focused on technical skills development and training, non-technical skills play a major role in creating a safe operating room environment.[39] Inadequate training in non-technical skills has been documented to result in increased technical skills errors, incorrect instrumentation and wrong site surgery.[40] Although non-technical skills apply in equal effect to both MAS and open surgery, they are strongly co-dependent on technical skills for optimal performance in the OR. Studies prove that non-technical skills are of major importance for surgical teams performing complex surgery in high technology operating rooms.[41]

ASSESSMENT OF TECHNICAL SKILLS

Assessment of technical skills is crucial in the evaluation and development skills in MAS training, as well in the motivation of the trainee.[42] Assessment methods are dependent on the complexity of the training modalities used and thus keep evolving based on the current developments of tools used. There are several modes of assessment which are classified as follows:

I Expert assessment

II Self-assessment

III Peer assessment

IV Equipment based assessment

EXPERT ASSESSMENT

Assessment and feedback are crucial in the development and progression of surgical skills. Several studies prove the importance of feedback, as feedback leads to greater retention and acquisition of skills.[43, 44] The general consensus is that technical skills assessment is a crucial part of surgical skills training.[45] The assessment however, for the major part remains subjective on the expert observers' part and has variable inter-rater validity or reliability.[46] Due to the complex nature of MAS training however, several tools have been developed for experts to objectively assess the trainees to the highest possible extent. These assessment tools are a combination of task deconstruction and procedure specific checklists. These checklists are hypothesized to " turn examiners into observers of behavior rather than interpreters of behavior, thereby removing the subjectivity of the evaluation process".[47] For example, a laparoscopic suturing exercise can be deconstructed into checklists such as needle loading, needle driving, pulling the suture through, knot tying technique, knot slippage and knot quality.[48]

The modes of expert feedback can be further classified into summative and formative feedback. Where summative feedback aims at credentialing and serves as an end tool of assessment, formative feedback aims at developing trainees' skills by giving structured and constructive feedback throughout the assessment process.[49]

SELF-ASSESSMENT

Similar to expert assessment, in self-assessment trainees are provided with the same tools to assess their performance as experts do. Research has proven that self-assessment is consistently similar to expert-assessment and plays a crucial role in self-regulation and professional development.[50–52] However, the effects of self-assessment are bi-fold. Self-assessment that is not administered well can lead to over assessment of one's performance and thus result in limited room for progression and potentially affect surgical outcomes. To achieve accurate self-assessment, several tools are used such as watching benchmark videos prior to performing a task, task demonstrations by experts prior to assessment and training on the usage of assessment forms.[53, 54]

PEER ASSESSMENT

Peer assessment fulfills a unique role in the assessment modes. In this form of assessment peers within a training group are required to evaluate one another using the same assessment tools that experts use. This has been proven to benefit not only the assessee but also the assessor.[55, 56] The resulting effects range from deep rather than surface learning and learning strategies from a different perspective.

Kruger and Dunnings' study further embolden these findings by showing that if people are unskilled in a particular domain, their incompetence deprives them of the metacognitive ability to realize it. But paradoxically, improving the skills of participants helped them to recognize their own limitations and abilities.[57]

EQUIPMENT BASED ASSESSMENT

Objective assessment has been the cornerstone of MAS skills assessment and can be evidenced by the immense research and tools developed for expert surgeons to be objective in their assessment of trainees. As previously mentioned, assessment by experts cannot truly be considered as objective even though the inter-rater reliability for most assessment tools are significantly high.

With the advent of technology, box trainers are now equipped with cameras for tracking, VR simulators are equipped with trackers and sensors to decode every aspect of performance into an objective measure. The most common amongst the metrics that can be extracted from these instruments are task duration and error in performance. However, these basic metrics cannot be used as a measure to determine the individual effort required to achieve the task or if learning objectives have been achieved.[58, 59]

Consequently, several other metrics have been identified and developed over time to accurately determine proficiency and performance monitoring metrics. These include trajectory, velocity and acceleration of instruments that can distinguish skills.[60, 61] These metrics are subject to change and advance as the training systems evolve and aim towards a simplistic and self-learning environment.

ASSESSMENT OF NON-TECHNICAL SKILLS

The assessment of non-technical skills unlike the technical skills assessment focuses on the interaction and ability of a surgeon to manage a team for the benefit of efficiency and patient safety. These skills are directly related to the technical skills of a surgeon. The first use case and assessment of non-technical skills was developed from the NOTECHS tool used in the aviation industry to train and measure teamwork and cognitive skills in the cockpit environment.[62]

The NOTECHS from the aviation industry has been successfully translated to Oxford NOTECHS and integrated into the operating room environment assessment and has been proven to address the key domains of non-technical skills such as leadership, teamwork, problem solving and situation awareness.[63] Similarly, several other tools such as the Observational teamwork assessment (OTAS) have been introduced with success.[64]

It is to be noted that unlike technical skills that can be translated universally amongst all surgeons, non-technical skills vary significantly pertaining to culture, organizational protocols, team and personality dynamics. Thus, the tool has to be modified based on these characteristics. With the introduction of head mounted displays in VR simulator, the training of non-technical skills can be integrated into a comprehensive learning program.

TRAINING CURRICULA IN MAS

Despite the evidence of effectiveness of simulators and training tools for MAS skills development, their availability and ease of access does not ensure educational effectiveness. In fact, there are several instances where these expensive equipment end up unused without proper curricula and expert faculty to administer and use them. The most efficient way of consolidating different stake holders in training in MAS is by means of effective curricula that encompasses all the aspects of training as described above. Such curricula are designed to include validated models of training and assessment.[65]

There are several curricula that have been designed ever since the advent of MAS recognizing the need for a radical and rapid adaptation for practicing and surgeons in practice. These range from grassroots courses limited to institutions to worldwide recognized courses. The Laparoscopic surgical skills (LSS) curriculum and the Fundamentals of Laparoscopic Surgery (FLS) are among the globally adopted curricula.[66, 67] Interestingly, as with surgical education that is regulated and endorsed by surgical communities and regulatory authorities, MAS training is far less regulated and mandatory in a recent survey.[68] The following chapters will focus on the current trends in training and the validation and future of training in MAS. The outline of this thesis is depicted in the following flow diagram for ease of reference.

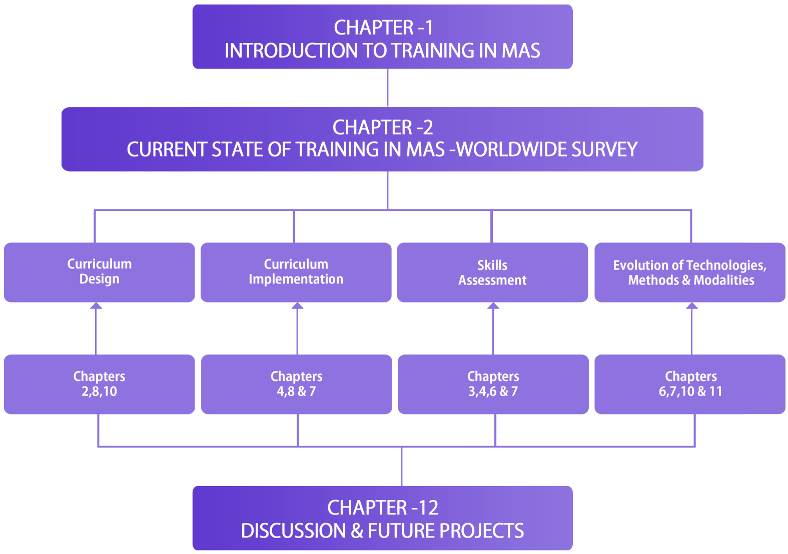


Figure 1.7: Outline of thesis

1

REFERENCES

[1] R. K. Reznick and H. MacRae, *Teaching surgical skills — changes in the wind,* New England Journal of Medicine **355**, 2664 (2006).

[2] *Laparoscopic surgery versus open surgery for colon cancer: short-term outcomes of a randomised trial,* The Lancet Oncology **6**, 477 (2005).

[3] V. Velanovich, *Laparoscopic vs open surgery,* Surgical Endoscopy **14**, 16 (2000).

[4] T. Jaschinski, C. G. Mosch, M. Eikermann, E. A. Neugebauer, and S. Sauerland, *Laparoscopic versus open surgery for suspected appendicitis,* Cochrane Database of Systematic Reviews (2018), 10.1002/14651858.cd001546.pub4.

[5] T. A. Gaskin, J. H. Isobe, J. L. Mathews, S. B. Winchester, and R. J. Smith, *Laparoscopy and the general surgeon,* Surgical Clinics of North America **71**, 1085 (1991).

[6] R. F. Martin, *Foreword,* Surgical Clinics of North America **90**, xiii (2010).

[7] M. P. Callery, D. B. Jones, L. Villegas, and B. E. Schneider, *Laparoscopic skills training,* Surgical Endoscopy **17**, 1879 (2003).

[8] D. J. Scott, P. C. Bergen, R. V. Rege, R. Laycock, S. T. Tesfay, R. Valentine, D. M. Euhus, D. Jeyarajah, W. M. Thompson, and D. B. Jones, *Laparoscopic training on bench models: better and more cost effective than operating room experience?* Journal of the American College of Surgeons **191**, 272 (2000).

[9] A. Pearson, A. Gallagher, J. Rosser, and R. Satava, *Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying,* Surgical Endoscopy And Other Interventional Techniques **16**, 130 (2001).

[10] E. Roy and T. Frank, *Spatial ability and learning the use of an angled laparoscope in a virtual environment,* Studies in Health Technology and Informatics **81**, 146–152 (2001).

[11] E. G. G. Verdaasdonk, L. P. S. Stassen, M. van der Elst, T. M. Karsten, and J. Dankelman, *Problems with technical equipment during laparoscopic surgery,* Surgical Endoscopy **21**, 275 (2007).

[12] S. M. D. Sørensen, M. M. Savran, L. Konge, and F. Bjerrum, *Three-dimensional versus two-dimensional vision in laparoscopy: a systematic review,* Surgical Endoscopy **30**, 11 (2015).

[13] P. Breedveld, H. G. Stassen, D. W. Meijer, and L. P. S. Stassen, *Theoretical background and conceptual solution for depth perception and eye-hand coordination problems in laparoscopic surgery,* Minimally Invasive Therapy & Allied Technologies **8**, 227 (1999).

[14] R. Berguer, D. L. Forkey, and W. D. Smith, *Ergonomic problems associated with laparoscopic surgery,* Surgical Endoscopy **13**, 466 (1999).

[15] A. Supe, G. Kulkarni, and K. Supe, *Ergonomics in laparoscopic surgery,* Journal of Minimal Access Surgery **6**, 31 (2010).

[16] L. Hull, S. Arora, R. Aggarwal, A. Darzi, C. Vincent, and N. Sevdalis, *The impact of nontechnical skills on technical performance in surgery: A systematic review,* Journal of the American College of Surgeons **214**, 214 (2012).

[17] A. Park, G. Lee, F. J. Seagull, N. Meenaghan, and D. Dexter, *Patients benefit while surgeons suffer: An impending epidemic,* Journal of the American College of Surgeons **210**, 306 (2010).

[18] F. M. Dimou, D. Eckelbarger, and T. S. Riall, *Surgeon burnout: A systematic review,* Journal of the American College of Surgeons **222**, 1230 (2016).

[19] J. D. Birkmeyer, J. F. Finks, A. O'Reilly, M. Oerline, A. M. Carlin, A. R. Nunn, J. Dimick, M. Banerjee, and N. J. Birkmeyer, *Surgical skill and complication rates after bariatric surgery,* New England Journal of Medicine **369**, 1434 (2013).

[20] S. Ganni, S. M. B. I. Botden, B. F. Hamilton, A. S. Bedi, D. Lomanto, B. R. G, and J. J. Jakimowicz, *Current state of training and evaluation of laparoscopic surgical skills,* Journal of Surgery: Open Access **3** (2017), 10.16966/2470-0991.148.

[21] J. Heemskerk, R. Zandbergen, J. G. Maessen, J. W. M. Greve, and N. D. Bouvy, *Advantages of advanced laparoscopic systems,* Surgical Endoscopy **20**, 730 (2006).

[22] K. R. Catchpole, A. E. Giddings, M. R. D. Leval, S. Gallivan, P. Godden, M. Utley, T. Dale, and G. Hirst, *Identifying and reducing systems failures through non-technical skills.* Royal College of Surgeons of Edinburgh Quincentenary Congress (2005).

[23] K. Subramonian, S. DeSylva, P. Bishai, P. Thompson, and G. Muir, *Acquiring surgical skills: A comparative study of open versus laparoscopic surgery,* European Urology **45**, 346 (2004).

1

[24] R. S. Haluck, R. L. Marshall, T. M. Krummel, and M. G. Melkonian, *Are surgery training programs ready for virtual reality? a survey of program directors in general surgery,* Journal of the American College of Surgeons **193**, 660 (2001).

[25] A. Cuschieri, N. Francis, J. Crosby, and G. B. Hanna, *What do master surgeons think of surgical competence and revalidation?* The American Journal of Surgery **182**, 110 (2001).

[26] R. Sidhu, E. Grober, L. Musselman, and R. Reznick, *Assessing competency in surgery: Where to begin?* Surgery **135**, 6 (2004).

[27] A. Darzi and S. Mackay, *Skills assessment of surgeons,* Surgery **131**, 121 (2002).

[28] A. G. Gallagher, E. M. Ritter, H. Champion, G. Higgins, M. P. Fried, G. Moses, C. D. Smith, and R. M. Satava, *Virtual reality simulation for the operating room,* Annals of Surgery **241**, 364 (2005).

[29] H. Schreuder, C. van den Berg, E. Hazebroek, R. Verheijen, and M. Schijven, *Laparoscopic skills training using inexpensive box trainers: which exercises to choose when constructing a validated training course,* BJOG: An International Journal of Obstetrics & Gynaecology **118**, 1576 (2011).

[30] S. M. Botden, S. N. Buzink, M. P. Schijven, and J. J. Jakimowicz, *Augmented versus virtual reality laparoscopic simulation: What is the difference?* World Journal of Surgery **31**, 764 (2007).

[31] M. K. Chmarra, N. H. Bakker, C. A. Grimbergen, and J. Dankelman, *TrEndo, a device for tracking minimally invasive surgical instruments in training setups,* Sensors and Actuators A: Physical **126**, 328 (2006).

[32] K. R. V. Sickle, D. A. M. III, A. G. Gallagher, and C. D. Smith, *Construct validation of the ProMIS simulator using a novel laparoscopic suturing task,* Surgical Endoscopy And Other Interventional Techniques **19**, 1227 (2005).

[33] V. Lahanas, C. Loukas, N. Smailis, and E. Georgiou, *A novel augmented reality simulator for skills assessment in minimal invasive surgery,* Surgical Endoscopy **29**, 2224 (2014).

[34] I. Choy and A. Okrainec, *Simulation in surgery: Perfecting the practice,* Surgical Clinics of North America **90**, 457 (2010).

[35] L. Panait, E. Akkary, R. L. Bell, K. E. Roberts, S. J. Dudrick, and A. J. Duffy, *The role of haptic feedback in laparoscopic simulation training,* Journal of Surgical Research **156**, 312 (2009).

[36] C. Våpenstad, E. F. Hofstad, T. Langø, R. Mårvik, and M. K. Chmarra, *Perceiving haptic feedback in virtual reality simulators,* Surgical Endoscopy **27**, 2391 (2013).

[37] O. A. J. van der Meijden and M. P. Schijven, *The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review,* Surgical Endoscopy **23**, 1180 (2009).

[38] E. G. G. Verdaasdonk, J. Dankelman, J. F. Lange, and L. P. S. Stassen, *Transfer validity of laparoscopic knot-tying training on a VR simulator to a realistic environment: A randomized controlled trial,* Surgical Endoscopy **22**, 1636 (2007).

[39] G. Fletcher, P. McGeorge, R. Flin, R. Glavin, and N. Maran, *The role of non-technical skills in anaesthesia: a review of current literature,* British Journal of Anaesthesia **88**, 418 (2002).

[40] C. C. Greenberg, S. E. Regenbogen, D. M. Studdert, S. R. Lipsitz, S. O. Rogers, M. J. Zinner, and A. A. Gawande, *Patterns of communication breakdowns resulting in injury to surgical patients,* Journal of the American College of Surgeons **204**, 533 (2007).

[41] A. Mishra, K. Catchpole, T. Dale, and P. McCulloch, *The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy,* Surgical Endoscopy **22**, 68 (2007).

[42] A. M. Derossis, G. M. Fried, M. Abrahamowicz, H. H. Sigman, J. S. Barkun, and J. L. Meakins, *Development of a model for training and evaluation of laparoscopic skills,* The American Journal of Surgery **175**, 482 (1998).

[43] T. Mahmood and A. Darzi, *The learning curve for a colonoscopy simulator in the absence of any feedback: No feedback, no learning,* Surgical Endoscopy **18**, 1224 (2004).

[44] M. C. Porte, G. Xeroulis, R. K. Reznick, and A. Dubrowski, *Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills,* The American Journal of Surgery **193**, 105 (2007).

1

[45] A. Cuschieri, N. Francis, J. Crosby, and G. B. Hanna, *What do master surgeons think of surgical competence and revalidation?* The American Journal of Surgery **182**, 110 (2001).

[46] R. K. Reznick, *Teaching and testing technical skills,* The American Journal of Surgery **165**, 358 (1993).

[47] G. Regehr, H. MacRae, R. K. Reznick, and D. Szalay, *Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination,* Academic Medicine **73**, 993 (1998).

[48] K. Moorthy, Y. Munz, A. Dosis, F. Bello, A. Chang, and A. Darzi, *Bimodal assessment of laparoscopic suturing skills,* Surgical Endoscopy **18**, 1608 (2004).

[49] Y. Watanabe, E. Bilgic, E. Lebedeva, K. M. McKendy, L. S. Feldman, G. M. Fried, and M. C. Vassiliou, *A systematic review of performance assessment tools for laparoscopic cholecystectomy,* Surgical Endoscopy **30**, 832 (2015).

[50] C. Rizan, J. Ansell, T. Tilston, N. Warren, and J. Torkington, *Are general surgeons able to accurately self-assess their level of technical skills?* The Annals of The Royal College of Surgeons of England **97**, 549 (2015).

[51] M. Ward, H. MacRae, C. Schlachta, J. Mamazza, E. Poulin, R. Reznick, and G. Regehr, *Resident self-assessment of operative performance,* The American Journal of Surgery **185**, 521 (2003).

[52] C. de Blacam, D. A. O'Keeffe, E. Nugent, E. Doherty, and O. Traynor, *Are residents accurate in their assessments of their own surgical skills?* The American Journal of Surgery **204**, 724 (2012).

[53] M. Ward, H. MacRae, C. Schlachta, J. Mamazza, E. Poulin, R. Reznick, and G. Regehr, *Resident self-assessment of operative performance,* The American Journal of Surgery **185**, 521 (2003).

[54] R. A. Stewart, L. S. Hauge, R. D. Stewart, R. L. Rosen, A. Charnot-Katsikas, and R. A. Prinz, *A CRASH course in procedural skills improves medical students' self-assessment of proficiency, confidence, and anxiety,* The American Journal of Surgery **193**, 771 (2007).

[55] S. Brown and P. Dove, *Opening mouths to change feet: some views on self and peer assessment,* in *Self and peer assessment, Standing Conference on Educational Development, Birmingham, SCED Paper*, Vol. 63 (1991).

[56] A. Brew, *Towards autonomous assessment: using self-assessment and peer assessment,* in *Assessment Matters In Higher Education (Society for Research Into Higher Education)* (Open University Press, 1999) Chap. 13, p. 159–71.

[57] J. Kruger and D. Dunning, *Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments.* Journal of Personality and Social Psychology **77**, 1121 (1999).

[58] D. Stefanidis, M. W. Scerbo, J. R. Korndorffer, and D. J. Scott, *Redefining simulator proficiency using automaticity theory,* The American Journal of Surgery **193**, 502 (2007).

[59] R. D. O'Donnell and F. T. Eggemeier, *Workload assessment methodology.* in *Handbook of perception and performance, cognitive processes and performance,* edited by K. R. Boff, L. Kaufman, and J. P. Thomas (Wiley New York, 1986) pp. 1–29.

[60] S. Yamaguchi, D. Yoshida, H. Kenmotsu, T. Yasunaga, K. Konishi, S. Ieiri, H. Nakashima, K. Tanoue, and M. Hashizume, *Objective assessment of laparoscopic suturing skills using a motion-tracking system,* Surgical Endoscopy **25**, 771 (2010).

[61] K. Moorthy, Y. Munz, A. Dosis, F. Bello, and A. Darzi, *Motion analysis in the training and assessment of minimally invasive surgery,* Minimally Invasive Therapy & Allied Technologies **12**, 137 (2003).

[62] R. Flin, K.-M. Goeters, H. Hoermann, R. Amalberti, C. Valot, and H. Nijhuis, *Development of the NOTECHS (non-technical) system for assessing pilots' CRM skills,* (2009) pp. 221–243.

[63] A. Mishra, K. Catchpole, and P. McCulloch, *The oxford NOTECHS system: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre,* Quality and Safety in Health Care **18**, 104 (2009).

[64] L. Hull, S. Arora, E. Kassab, R. Kneebone, and N. Sevdalis, *Observational teamwork assessment for surgery: Content validation and tool refinement,* Journal of the American College of Surgeons **212**, 234 (2011).

[65] J. Dent, *A Practical Guide for Medical Teachers* (Elsevier, 2017).

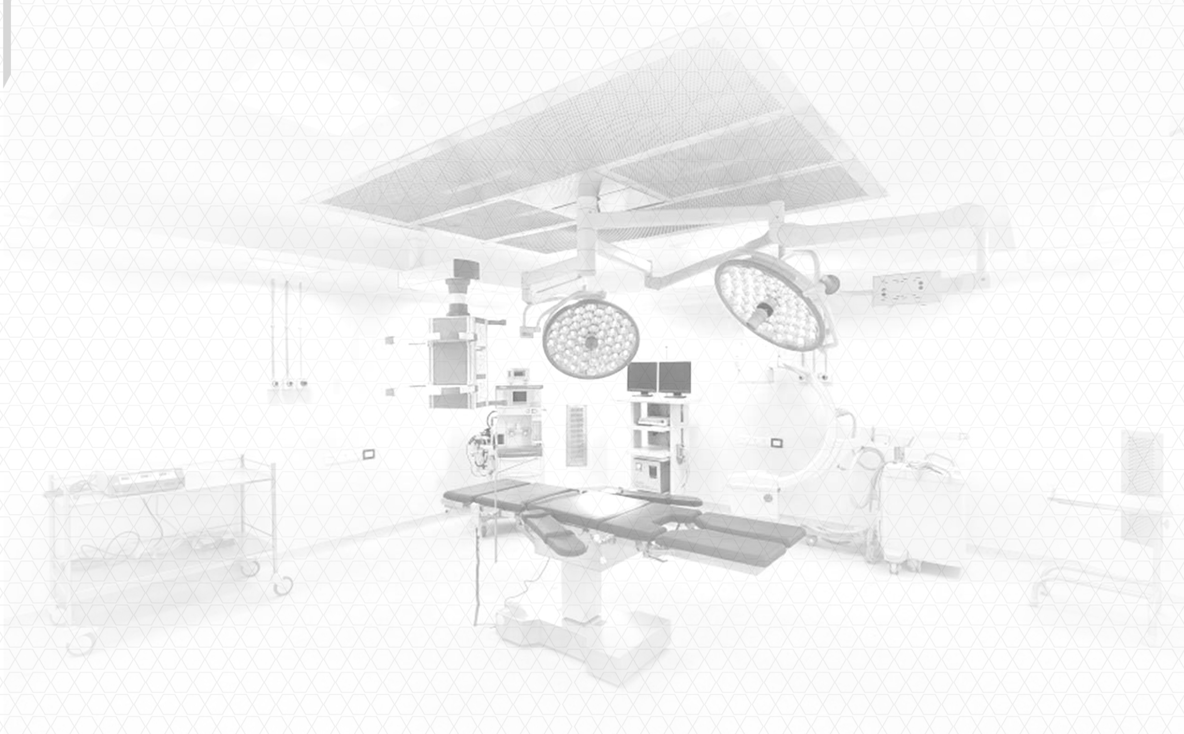
[66] J. J. Jakimowicz and S. Buzink, *Training curriculum in minimal access surgery,* in *Training in Minimal Access Surgery* (Springer London, 2015) pp. 15–34.

1

[67] G. Sroka, L. S. Feldman, M. C. Vassiliou, P. A. Kaneva, R. Fayez, and G. M. Fried, *Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial,* The American Journal of Surgery **199**, 115 (2010).

[68] S. Ganni, S. M. B. I. Botden, B. F. Hamilton, A. S. Bedi, D. Lomanto, B. R. G, and J. J. Jakimowicz, *Current state of training and evaluation of laparoscopic surgical skills,* Journal of Surgery: Open Access **3** (2017), 10.16966/2470-0991.148.

2



CURRENT STATE OF TRAINING AND EVALUATION OF LAPAROSCOPIC SURGICAL SKILLS

S. Ganni, S.M.B.I. Botden, B.F. Hamilton, A.S. Bedi, D. Lomanto, J.J. Jakimowicz

Published in Journal of Surgery: Open access, 3(3), March 2017

2

ABSTRACT

Objective The aim of this study is to understand the current state of training practices and evaluation in laparoscopic surgery in a global context.

Design An open-ended three part questionnaire was designed to gather the opinions about the current state of, adequacy of, and the need for a standard in laparoscopic surgical training.

Participants Members of the European Association for Endoscopic Surgery (EAES), Endoscopic and Laparoscopic Surgeons of Asia (ELSA) and Association of Surgeons of India (ASI) were asked to participate in the survey.

Results Of the 663 responses received, 83.6% were surgeons (64.6% in a teaching position) and 12.6% were surgical residents in training. Most respondents (75.4%) had performed over 200 laparoscopic procedures. Most (72.1%) training programs were approved/endorsed by local surgical associations or government health authorities and of the courses taught by surgical associations the majority had certified trainers (71.1%). In lower Human Development Index (HDI) countries significantly less courses are taught by certified trainers (68.2% versus 54.6%, $p < 0.001$). Only 26.8% stated that their respective government health authorities participated in the certification of laparoscopic surgery; certification was considered important by 63.6%. However, only 17.8% of government health authorities contributed to ensure the quality of laparoscopic training, mostly in very high HDI. Only 3.3% of respondents considered the laparoscopic training and education in their country to be optimal and 51.9% rated it insufficient. Most respondents (86.3%) stated that there is a need for the standardization of laparoscopic training and 88.3% stated that standardization of laparoscopic training is important.

Conclusion Regardless of demographic and experience factors, there was a general consensus that that there is a need for standardisation in mandatory training of laparoscopic surgical skills, although currently not obligatory in most countries.

Introduction

Minimal-access surgery (MAS) has revolutionized the field of surgery over the past few decades; it is now considered to be the gold-standard for many surgical procedures, due to the numerous benefits it offers to patients [1–3]. Laparoscopic surgery is more complex than open surgery and requires a new set of skills that are different from conventional surgery. The surgeon has to become proficient in handling the new instruments, the considerable loss of haptic feedback, dealing with the counter-intuitive manipulation of the instruments, eye-hand coordination and the two-dimensional representation of the three-dimensional operating field [4–6].

Multiple models have been developed to train laparoscopic skills via simulation, including box trainers, animal models, virtual reality (VR) and augmented reality (AR) simulators.[7–10] The skills acquired by using simulator training have been proven by numerous studies, which show an effective transfer to the operating theatre [11–14].

However, despite this strong evidence that proves the efficacy of training, several reports state the underutilization of simulation and lack of integration within the standard surgical residency training programs [15–17]. In addition, Chang et al, report that simulation based training curricula should be a mandatory part of the residency curricula [18].

With the implementation of restrictive working hours in both Europe and North America, surgical residents have reduced laparoscopic surgery exposure. This has been shown to reduce their experience and surgical skills [19–21]. Which creates an even greater need for an obligatory structured training curriculum, to ensure the quality of all laparoscopic surgical procedures globally. One of the challenges in implementing novel training curricula are the different attitudes and perceptions towards training among the surgical community including surgeons, residents, health authorities and the surgical industry [22]. Therefore, not only the availability of training resources, but also the initiatives of local and national health authorities and the surgical industry play an important role in delivering standard training practices [23, 24].

The objective of this study was to address the influencing factors on laparoscopic training and to determine the current training practices in laparoscopic surgery globally. Also, the opinion on the quality of the current training and need for standardisation was assessed.

Materials and Methods

This study was designed to evaluate the opinions of experienced laparoscopic surgeons and surgical residents, who expressed an interest in laparoscopic techniques, regarding the current state of training and standardization of laparoscopic training.

Subjects

All members of the European Association for Endoscopic Surgeons (EAES), Endoscopic and Laparoscopic Surgeons of Asia (ELSA) and The Association of Surgeons of India (ASI) were contacted by email to complete the questionnaire. The members included both experienced surgeons and inexperienced surgical residents from around the world and there was an 10.1% response rate. They were asked to complete the questionnaire, with a reminder email sent two weeks after the initial email. The participants were allotted to two groups based on their experience (more than 200 laparoscopic procedures or less than 200 laparoscopic procedures). The participants asked were from 73 countries, which were also divided in groups using the Human Development Index (HDI) ranking, because both developed and developing countries were included in this study. The countries were ranked Higher HDI and Lower HDI based on the United Nations Development Programme ranking list [25]. The first fifty countries in the list were considered as Very High and High HDI, whereas the remainder were considered Medium and Low HDI countries.

Questionnaire

The questionnaire (Appendix A) was designed to address demographic differences in perspectives depending on country, grouped by Human Development Index; surgical experience (surgeon or resident and number of procedures performed) and gender. The details of current training practices in their respective countries were evaluated along

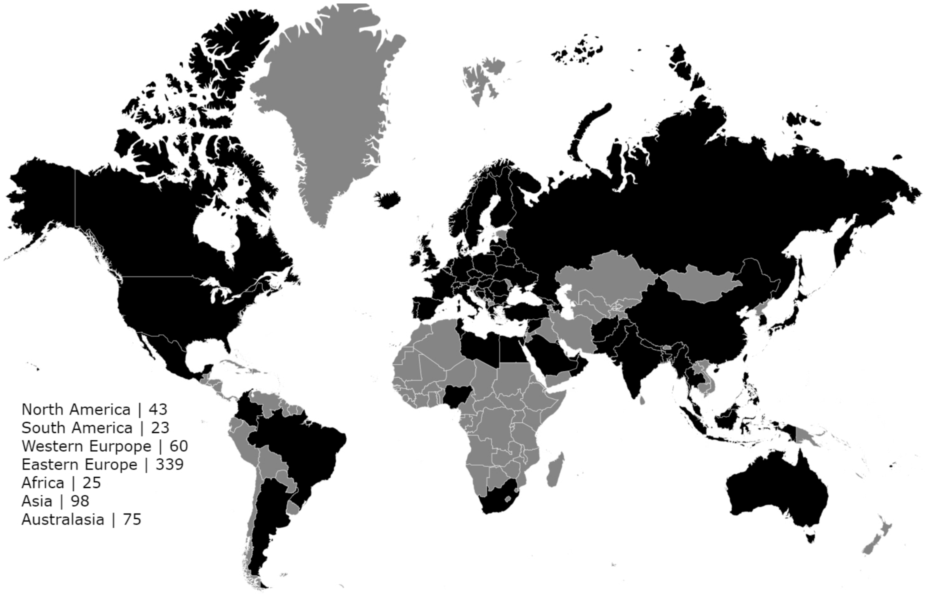


Figure 2.1: Map depicting the countries from which responses were received with continent wise segregation. Black are countries from which a response was received.

with their facilitation and implementation. The questionnaire ended with a statement on the need for standardisation for laparoscopic training and an open-ended question for general remarks.

Statistical Analysis

All data were processed, coded and analysed using SPSS (Version 22, IBM Corp.). Because the data are nonparametric, the Mann-Whitney U and Chi-square tests were conducted to determine the statistical differences between the responses to individual questions. A p-value of < 0.05 was considered a statistically significant difference. Where percentages sum to a total above 100% this is because several questions were multiple-choice and respondents responded with more than one answer. Such percentages are indicated by a $'+'$.

Results

Demographics

A total of 663 responses were obtained from 73 different countries around the world (Figure 2.1). The countries were divided in two groups based on their Human

	Question	Lower HDI Group (%)	Very High HDI Group (%)	Pearson Chi-Square p-value
1	**Indicate your experience level**			
	Resident in training	9.9	13.9	
	Surgeon	80.3	85.2	<0.001*
	Other	9.9	0.9	
2	**Training of MAS in your country is**			
	Courses as part of residency curriculum (obligatory)	45.3	47.1	0.672
	Through special courses (non-obligatory)	66.5	54.2	0.003*
	Other	3.9	3.0	0.536
3	**Which type of courses are organized?**			
	Mixed	82.6	85.6	
	Hands-on	13.4	9.8	0.446
	Theoretical	4.1	4.6	
4	**What is the level of the courses organized?**			
	Basic	28.6	13.5	<0.001
	Advanced	8.4	10.1	0.484
	Both	58.1	63.7	0.176
	Speciality	26.6	29.7	0.419
	Other	3.0	2.8	0.909
5	**Which type of training do the courses involve?**			
	Box Trainers	74.4	65.8	0.028
	Tissue Models	44.8	44.7	0.982
	Artificial models	19.2	20.2	0.765
	Live Animals	53.7	48.4	0.207
	Cadaver training	13.8	23.9	0.003*
	VR training	18.7	22.4	0.289
	Augmented reality training	4.4	3.9	0.734
	Simulator	38.4	39.1	0.861
6	**Are the trainers in the courses certified?**			
	Yes	68.2	54.6	
	No	17.1	34.3	<0.001*
	Yes by whom	14.7	10.9	
7	**Does your country require certification in Laparoscopic surgery?**			
	Yes	48.8	11.2	
	No	42.5	87.1	<0.001*
	Yes with why	8.8	1.7	
8	**Do you think training and education for laparoscopic surgery in your country is currently**			
	Insufficient	65.0	45.5	
	Sufficient	13.8	24.5	<0.001*
	Adequate	16.9	27.4	
	Optimal	4.4	56.3	
9	**Do you think the need for standardisation in laparoscopic surgical training exists?**			
	Yes	87.6	85.5	
	No	10.6	13.3	0.573
	Yes with why	1.9	1.2	

Table 2.1: Chi-squared tests between the High HDI and Lower HDI groups for questions deemed crucial in the survey.

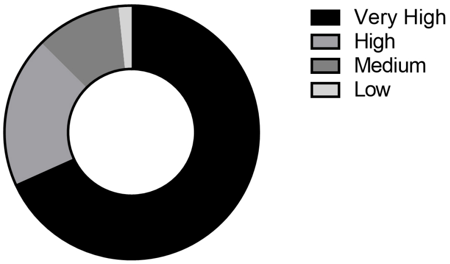


Figure 2.2: The distribution of HDI between the participants in the survey.

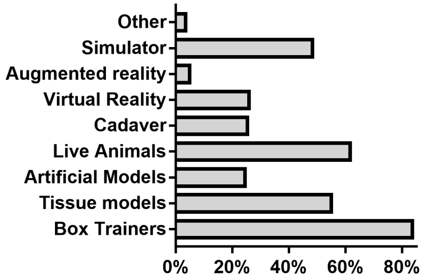


Figure 2.3: Different modalities used in education and training MAS.

Development Indicators ranking, in which the first group had a Very High HDI (developed country, 49 countries, $n = 460$ participants), and the remaining were in the second group, those countries with High, Medium or Low HDI (26 countries, $n = 202$ participants) (Figure 2.2). Chi-squared tests between the Very High HDI and Lower HDI groups for questions deemed crucial have been calculated and are shown for completeness in Table 2.1.

Of all participants, 83.6% were surgeons and 12.6% were surgical residents in training ($n = 641$). Most participants were male (87.3%), with no significant different between the HDI groups or surgeons/residents. Of the surgeons, 64.6%+ were also in a teaching position, while the remainder performed the procedures themselves.

When dividing the participants based on laparoscopic experience, the majority (75.4%) performed more than 200 laparoscopic procedures and were marked as experienced surgeons for this study. The remainder were marked as less experienced.

CURRENT LAPAROSCOPIC TRAINING

The vast majority of training courses are organized by either teaching hospitals (46.9%+), University hospitals (57.8%+) or surgical associations (66.2%+). However, 40.9%+ of respondents suggested that courses are organized by the surgical industry. Most of the training programs were approved/endorsed by local surgical associations or government health authorities (72.1%). Most of the courses taught by surgical associations had certified trainers (71.1%). However, there were significantly less courses taught by certified trainers in lower HDI countries (68.2%+ versus 54.6%+, $p < 0.001$). Although 69.3%+ responded that there are special courses for laparoscopic training, only 56.4%+ stated that is was part of the residency curriculum.

The majority of the responses stated that the courses used both theory (instructions and lectures) and hands-on training (84.7%), while 10.7% had only hands-on and 4.6% only theoretical courses available in their country. The majority of training modalities used during the courses were box trainers (83.6%+), live animals (61.6%+), tissue models (54.9%+) and simulator training (48.3%+) (Figure 2.3). The majority, of the courses were multimodal (78.9%+), of which most often both the box trainer and VR/AR simulators were used (45.2%+). When live animals were used, this was most often a unimodal training course (64%+).

Most courses included indicator procedures as a base for the training, for which laparoscopic cholecystectomy was used most often (93.0%+), followed by appendectomy (75.2%+). These procedures were included as procedural tasks in 75.6%+ of the courses, with the remainder only using component tasks. Live surgery demonstrations featured in 80.0% of the courses, however, in countries with a lower HDI rating, a significantly lower proportion of courses had live surgery demonstrations ($U = -3.713$, $p < 0.001$). In 75.2% of the participants both basic and advanced courses were provided, which was mainly in the very high HDI countries (28.6% versus 13.5%, $p < 0.001$).

One third responded that there was no assessment of the course, or only partly and particularly the knowledge component was not assessed (30%). The assessment tools used in the courses were validated tools such as Competency Assessment Tool (CAT), Observational Clinical Human Reliability Index (OCHRA), McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) and (Objective Structured assessment of technical skill) OSATS. Other responses on the assessment were "occasionally theoretical examination", "there is no special system of evaluation" and "individual interview".

OPINION ON CERTIFICATION AND STANDARDIZATION

Only 26.8% stated that their respective government health authorities participated in the certification of laparoscopic surgery with no difference between the HDI-grouped data with a U leading to (often significantly more than) $p > 0.05$ in each case. In 52 of 73 countries there is currently no certification in laparoscopic surgery required during the surgical training. Furthermore, certification was considered important by 63.6%, with the most stated rational legal purposes (Figure 2.4). Only 3.3% of respondents considered the laparoscopic training and education in their country to be optimal and 51.9% considered the training insufficient.

The vast majority of respondents (86.3%) were of the opinion that there is a need for standardization of laparoscopic training. Experienced surgeons more often stated a

2

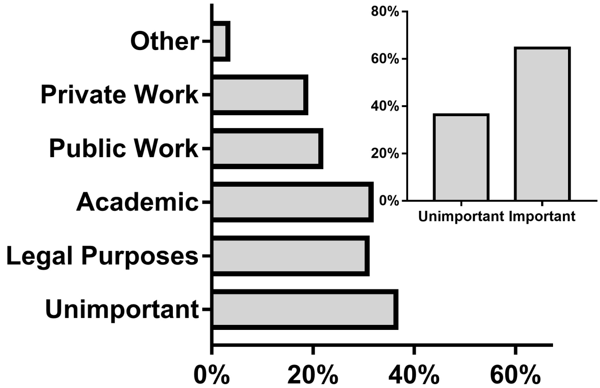


Figure 2.4: Reasons for importance of certification.

need for standardization (16.7% versus 8.1% $p = 0.0147$). On the contrary, according to the participants in only 17.8% the government health authorities contributed to ensure the quality of laparoscopic training, which was in 14 mainly Very High HDI countries (10). Less than half of the respondents were familiar with either the Fundamentals in Laparoscopic Surgery course (FLS) (47.4%) and Laparoscopic Surgical Skills curriculum (LSS) (45.33%) in their country. However, still 88.3% stated that standardization of laparoscopic training is important.

Discussion

This study evaluates the current laparoscopic training and certification of 73 countries in both developed and developing countries. While similar studies have been conducted to assess the same areas, the scope of this study is wider, because it goes beyond the limitations of a country- or continent-specific study [26–29]. With 663 participants, this large sample size represented both experienced and inexperienced laparoscopic surgeons and surgical residents from countries ranging from a very high HDI to low HDI. The overall high level of experience in both practice and teaching benefits the results of this study. However, the opinions of the experienced surgeons were not significantly different from the results of the less experienced responders.

A review by Pellegrini et al shows that integration of laparoscopic skills training within the residency training programs shows marked improvement in the development and retention of skills [30]. However, this study indicated that the majority of laparoscopic skills training are still conducted through non-obligatory courses outside the surgical residency training. Furthermore, some respondents in the current study indicated the nature of these courses as "if you choose to pay, you can do the course", which makes the threshold for lower income surgeons and surgical residents much higher [31].

Most courses involve box trainers, tissue models, live animals, and (VR) simulator training. Often multimodality training was used and both component and procedural tasks were trained in these courses. The majority of courses contained both a theoretical and hands-on part of the training. This shows an overall international consensus in the best training practices for laparoscopic surgery. Although many courses included live surgery component, there was a significantly lower proportion of courses with live surgery in the lower HDI country group. This could be attributed to the high costs associated with the setup of a wet-lab and with anaesthetising animals [15, 32].

Although several validated assessment methods were used, one third stated that there was no assessment of the courses at all. Interestingly, the majority of courses without an assessment were organized in Lower HDI countries.

The majority of participants indicated a need for standardized training curriculum even though one third of countries do not currently have certification of laparoscopic skills. Respondents were proactive in signing up to receive information on existing internationally accredited curricula such as the Fundamentals of Laparoscopic Surgery (FLS) and the Laparoscopic Surgical Skills (LSS) curriculum supports the thesis that there is a universal desire for standardisation in laparoscopic training.

Some of the limitations of this study surround its design; multiple-choice questions were asked in order that the complex range of respondents and opinions could be collated accurately; this makes some of the statistics more difficult to interpret however, it has yielded a more broad overview of the data. In retrospect, some of the questions in this study could have benefited from using a more objective measure, such as a Likert scale. Further, in order to attain a global range of responses, three organisations were used to distribute the survey. This may have resulted in a bias towards more experienced surgeons (those more likely to belong to a surgical association) but this was taken into account by separating the results for more and less experienced surgeons.

CONCLUSION

There appears to be an overall conformity in usage of methods and methodologies in training laparoscopic skills but the use of standard training curriculum, even within geographical regions and countries, is lacking. However, regardless of demographic and experience factors, there was a general consensus that there is a need for standardisation in mandatory training of laparoscopic surgical skills, although currently not obligatory in most countries. Given the state of advances and infiltration of laparoscopic surgery and allied technology in global practice and with the increasing restrictive working hours there is a need for a standard in training and evaluation of laparoscopic surgical skills.

2

REFERENCES

[1] K. Hida, S. Hasegawa, Y. Kinjo, K. Yoshimura, M. Inomata, M. Ito, Y. Fukunaga, A. Kanazawa, H. Idani, Y. Sakai, and M. Watanabe, *Open versus laparoscopic resection of primary tumor for incurable stage IV colorectal cancer,* Annals of Surgery **255**, 929 (2012).

[2] N. Byrge, R. G. Barton, T. M. Enniss, and R. Nirula, *Laparoscopic versus open repair of perforated gastroduodenal ulcer: a national surgical quality improvement program analysis,* The American Journal of Surgery **206**, 957 (2013).

[3] A. van Dijk, M. Lamberts, C. van Laarhoven, J. Drenth, M. Boermeester, and P. de Reuver, *Laparoscopy in cholecysto-choledocholithiasis,* Best Practice & Research Clinical Gastroenterology **28**, 195 (2014).

[4] D. J. Scott, P. C. Bergen, R. V. Rege, R. Laycock, S. T. Tesfay, R. Valentine, D. M. Euhus, D. Jeyarajah, W. M. Thompson, and D. B. Jones, *Laparoscopic training on bench models: better and more cost effective than operating room experience?* Journal of the American College of Surgeons **191**, 272 (2000).

[5] A. Pearson, A. Gallagher, J. Rosser, and R. Satava, *Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying,* Surgical Endoscopy And Other Interventional Techniques **16**, 130 (2001).

[6] E. Roy and T. Frank, *Spatial ability and learning the use of an angled laparoscope in a virtual environment,* Studies in Health Technology and Informatics **81**, 146–152 (2001).

[7] A. M. Derossis, G. M. Fried, M. Abrahamowicz, H. H. Sigman, J. S. Barkun, and J. L. Meakins, *Development of a model for training and evaluation of laparoscopic skills,* The American Journal of Surgery **175**, 482 (1998).

[8] S. M. B. I. Botden, S. N. Buzink, M. P. Schijven, and J. J. Jakimowicz, *ProMIS augmented reality training of laparoscopic procedures face validity,* Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare **3**, 97 (2008).

[9] S. M. Botden, S. N. Buzink, M. P. Schijven, and J. J. Jakimowicz, *Augmented versus virtual reality laparoscopic simulation: What is the difference?* World Journal of Surgery **31**, 764 (2007).

[10] S. M. B. I. Botden and J. J. Jakimowicz, *What is going on in augmented reality simulation in laparoscopic surgery?* Surgical Endoscopy **23**, 1693 (2008).

[11] T. Mori, N. Hatano, S. Maruyama, and Y. Atomi, *Significance of "hands-on training" in laparoscopic surgery,* Surgical Endoscopy **12**, 256 (1998).

[12] J. R. Korndorffer, J. B. Dunne, R. Sierra, D. Stefanidis, C. L. Touchard, and D. J. Scott, *Simulator training for laparoscopic suturing using performance goals translates to the operating room,* Journal of the American College of Surgeons **201**, 23 (2005).

[13] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, and R. M. Satava, *Virtual reality training improves operating room performance,* Annals of Surgery **236**, 458 (2002).

[14] M. P. Schijven, J. J. Jakimowicz, I. A. M. J. Broeders, and L. N. L. Tseng, *The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum,* Surgical Endoscopy **19**, 1220 (2005).

[15] A. Park and D. B. Witzke, *Training and educational approaches to minimally invasive surgery: State of the art,* Seminars in Laparoscopic Surgery **9**, 198 (2002).

[16] W. Kolkman, M. A. J. V. de Put, W. B. V. den Hout, J. B. M. Z. Trimbos, and F. W. Jansen, *Implementation of the laparoscopic simulator in a gynecological residency curriculum,* Surgical Endoscopy **21**, 1363 (2006).

[17] B. Dunkin, G. L. Adrales, K. Apelgren, and J. D. Mellinger, *Surgical simulation: a current review,* Surgical Endoscopy **21**, 357 (2006).

[18] L. Chang, J. Petros, D. T. Hess, C. Rotondi, and T. J. Babineau, *Integrating simulation into a surgical residency program,* Surgical Endoscopy **21**, 418 (2006).

[19] K. J. Breen, A. M. Hogan, and K. Mealy, *The detrimental impact of the implementation of the european working time directive (EWTD) on surgical senior house officer (SHO) operative experience,* Irish Journal of Medical Science **182**, 383 (2013).

[20] T. H. Kamine, S. Gondek, and T. S. Kent, *Decrease in junior resident case volume after 2011 ACGME work hours,* Journal of Surgical Education **71**, e59 (2014).

[21] N. Ahmed, K. S. Devitt, I. Keshet, J. Spicer, K. Imrie, L. Feldman, J. Cools-Lartigue, A. Kayssi, N. Lipsman, M. Elmi, A. V. Kulkarni, C. Parshuram, T. Mainprize, R. J. Warren, P. Fata, M. S. Gorman, S. Feinberg, and J. Rutka, *A systematic review of the effects of resident duty hour restrictions in surgery,* Annals of Surgery **259**, 1041 (2014).

[22] D. Stefanidis, C. E. Acker, D. Swiderski, B. T. Heniford, and F. L. Greene, *Challenges during the implementation of a laparoscopic skills curriculum in a busy general surgery residency program,* Journal of Surgical Education **65**, 4 (2008).

[23] R. Kneebone, *Simulation in surgical training: educational issues and practical implications,* Medical Education **37**, 267 (2003).

[24] B. Zendejas, R. Brydges, S. J. Hamstra, and D. A. Cook, *State of the evidence on simulation-based training for laparoscopic surgery,* Annals of Surgery **257**, 586 (2013).

[25] UNDP, *Human development report,* (2015).

[26] M. P. Schijven, J. T. M. Berlage, and J. J. Jakimowicz, *Minimal-access surgery training in the netherlands: a survey among residents-in-training for general surgery,* Surgical Endoscopy **18**, 1805 (2004).

[27] D. Rattner, K. Apelgren, and W. Eubanks, *The need for training opportunities in advanced laparoscopic surgery,* Surgical Endoscopy **15**, 1066 (2001).

[28] H. Yeo, K. Viola, D. Berg, Z. Lin, M. Nunez-Smith, C. Cammann, R. H. Bell, J. A. Sosa, H. M. Krumholz, and L. A. Curry, *Attitudes, training experiences, and professional expectations of US general surgery residents,* JAMA **302**, 1301 (2009).

[29] V. N. Palter, N. Orzech, R. Aggarwal, A. Okrainec, and T. P. Grantcharov, *Resident perceptions of advanced laparoscopic skills training,* Surgical Endoscopy **24**, 2830 (2010).

[30] C. A. Pellegrini, A. L. Warshaw, and H. T. Debas, *Residency training in surgery in the 21st century: A new paradigm,* Surgery **136**, 953 (2004).

[31] I. Choy, S. Kitto, N. Adu-Aryee, and A. Okrainec, *Barriers to the uptake of laparoscopic surgery in a lower-middle-income country,* Surgical Endoscopy **27**, 4009 (2013).

[32] M. H. Bernstein, *Is justification of animal research necessary?* JAMA: The Journal of the American Medical Association **269**, 1113 (1993).

3

sum

SELF-ASSESSMENT IN LAPAROSCOPIC SURGICAL SKILLS TRAINING: IS IT RELIABLE?

S.Ganni, M.K.Chmarra, R.H.M.Goossens, J.J.Jakimowicz

Published in Surgical Endoscopy, 31(6), 2451-2456, September 2016

3

Abstract

Background

The concept of self-assessment has been widely acclaimed for its role in the professional development cycle and self-regulation. In the field of medical education, self-assessment has been most used to evaluate the cognitive knowledge of students. The complexity of training and evaluation in laparoscopic surgery has previously acted as a barrier in determining the benefits self- assessment has to offer in comparison with other fields of medical education.

Methods

Thirty-five surgical residents who attended the 2-day Laparoscopic Surgical Skills Grade 1 Level 1 curriculum were invited to participate from The Netherlands, India and Romania. The competency assessment tool (CAT) for laparoscopic cholecystectomy was used for self- and expert-assessment and the resulting distributions assessed.

Results

A comparison between the expert- and self-assessed aggregates of scores from the CAT agreed with previous studies. Uniquely to this study, the aggregates of individual sub-categories — 'use of instruments'; 'tissue handling'; and errors 'within the component tasks' and the 'end product?' from both self- and expert-assessments —

were investigated. There was strong positive correlation ($r_s > 0.5$; $p < 0.001$) between the expert- and self-assessment in all categories with only the 'tissue handling' having a weaker correlation ($r_s = 0.3$; $p = 0.04$). The distribution of the mean of the differences between self-assessment and expert-assessment suggested no significant difference between the scores of experts and the residents in all categories except the 'end product' evaluation where the difference was significant ($W = 119$; $p = 0.03$).

CONCLUSION

Self-assessment using the CAT form gives results that are consistently not different from expert- assessment when assessing one's proficiency in surgical skills. Areas where there was less agreement could be explained by variations in the level of training and understanding of the assessment criteria.

INTRODUCTION

The concept of self-assessment has been widely acclaimed for its role in professional development cycle and self- regulation [1, 2]. The term self-assessment itself, however, is loosely defined and is thus the subject of criticism regarding its effectiveness in practice [3]. There has been considerable debate as to the efficacy of self-assessment but most criticism of self-assessment concerns the methodologies used, rather than the pedagogy itself [4–6]. Several educational psychology studies assert that self- assessment should be integrated from within the training phase to inculcate it as a lifelong professional habit [7–9]. In professional practice, however, the reality is that self- assessment is most commonly used as an evaluative tool for final performance [2].

In the field of medical education, self-assessment is mostly used to evaluate the cognitive knowledge of students [10, 11]. In surgical training, where acquisition of complex surgical skills such as cognitive, psychomotor and decision-making skills is required, self-assessment has not gained enough attention. In laparoscopic surgery, assessment of surgical skills is done either by surgical experts or by means of virtual reality (VR) simulators [12, 13]. Though VR simulators offer a certain degree of self-assessment, it is limited to psychomotor skills assessment against pre-defined benchmarks [14].

In addition to the complexity of assessment of skills in laparoscopic surgery, the costs — in terms of actual hours and time spent away from the operating theatre — of training and evaluating surgical residents by expert are very high [15]. An effective self-assessment tool could help in reflection on performance and assessment of trainees in the course of training and thus sequentially reducing the workload of expert surgeons.

The aim of this study was to assess the validity of using self-assessment within the Laparoscopic Surgical Skills curriculum (an initiative of the European Association of Endoscopic Surgery) [16]. The competency assessment tool (CAT) for laparoscopic cholecystectomy (LC) was used for self-assessment and expert-assessment in this study, and the results were compared.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-five surgical residents who attended the 2-day Laparoscopic surgical skills Grade 1 Level 1 curriculum were invited to participate (Table 3.2). Their expertise level ranged from PGY-2 to PGY-3. All of the surgical residents had prior experience using both box trainers and VR simulators.

All participants voluntarily enrolled in the study and signed an informed consent prior to the start of the curriculum. They also had to fill in a demographic questionnaire with data pertaining to experience in laparoscopic surgery and time spent preparing for the curriculum.

Six expert surgeons from the respective locations conducting the curriculum were invited to participate as expert assessors. Their experience in laparoscopic surgery ranged from 5 to 25 years, each with more than 200 laparoscopic procedures performed as a main surgeon. They also all had experience using the CAT form as a form of evaluation previously.

TASK

The participants had to fill out a multiple choice questionnaire on the basics of laparoscopic surgery to be admitted into the curriculum. During the curriculum, they participated in interactive discussions on the basics of laparoscopic surgery and LC, training on VR simulators and box trainers.

Each participant performed an LC procedure on a pig liver placed in a box trainer. The box trainer with ports that mimicked incision points was placed on a height adjustable table with monitors and equipment in place. Each participant was assisted by a fellow participant, who held the camera and, when needed, the instruments: playing the role of an assistant. The expert surgeons instructed the participants on the procedural tasks prior to the procedure and intervened whenever they deemed instruction was necessary. However, the assessors were asked not to express their opinions on the performance whilst the participants performed the procedure. After completing the procedure, both the participants and expert surgeons had to fill in the CAT form independently of one another.

ASSESSMENT

The CAT form was used in the study for self-assessment and expert-assessment. The CAT is an operation-specific assessment tool that was adapted for the LC procedure for use within the curriculum [17]. The evaluation criteria are spread across three procedural tasks: exposure of cystic artery and cystic duct, cystic pedicle dissection and resection of gallbladder from the liver. Within these tasks, the performance was rated on a five-point task-specific scale based on the usage of instruments, handling of tissue with the non-dominant hand (NDH), errors within each task and the end product of each task.

STATISTICAL ANALYSIS

Analysis was done comparing the expert- and self-assessment scores based on the above-mentioned criteria within the tasks. Scores for each category were summed to form aggregate scores for each, related, category. The scores for all the criteria were also calculated in order to compare our results with other studies. Obtained data were analysed using GraphPad Prism (Version 7.00). Spearman's rank correlation was used to assess the correlation between the expert- and self-assessment results. The Wilcoxon matched-pairs signed-rank test was used to assess whether the population mean ranks differ. A p value of < 0.05 was considered statistically significant.

RESULTS

CORRELATION IS SEEN BETWEEN EXPERT- AND SELF- ASSESSMENT

Figures 3.1 and 3.2 show exemplar scatter plots for the aggregate scores of the all criteria and tissue-handling data, respectively. There is statistically significant positive correlation between self-assessed answers and expert's opinions. All groupings show a Spearman's rank of greater than 0.5, corresponding to a strong positive correlation with the exception of the tissue handling and usage of NDH grouping which shows a weaker positive correlation of 0.3042.

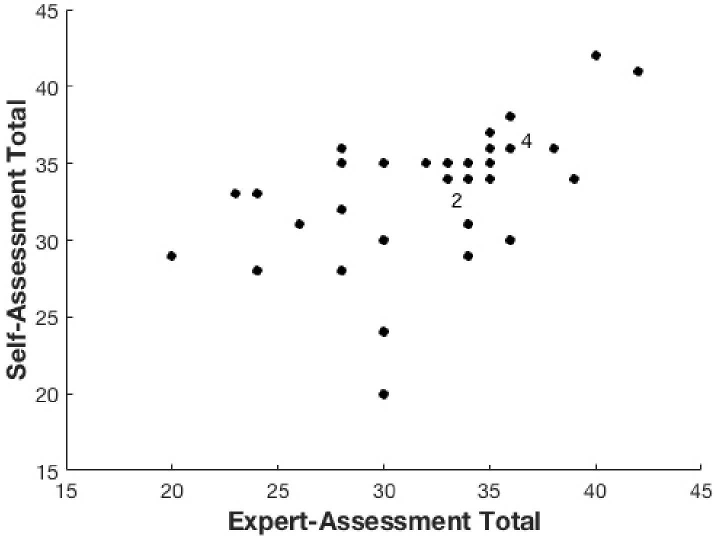


Figure 3.1: Self-assessment (SA) versus expert-assessment (EA) score for aggregated responses to all questions. Numbers to the right of data points show the number of coincident data points at the same coordinates, i.e., the number of people with the same combination of SA and EA scores

SIMILAR DISTRIBUTION OF RESPONSES BETWEEN EXPERT- AND SELF-ASSESSMENT

The statistics calculated to compare their distribution are shown in Table 3.1. Figure 3.3 shows the distribution of the responses of both experts and participants. Figure 3.4 demonstrates how similar the means (±SEM) of the grouped, aggregated data are. With the exception of the 'end product evaluation' criterion, all the groupings result in a p value greater than the 0.05 threshold for rejecting the null hypothesis. The 'end product evaluation' criterion has a Wilcoxon $p-$value of 0.0339 which suggests that in the case of the 'end-product evaluation' criterion a difference in the distribution of the mean difference was seen. Furthermore, there was no significant difference between the mean of the differences in scores for men (1.17, SD =3.32; SEM 0.76) and women (0.94, SD = 5.60; SEM 1.37) whose demographic distribution can be seen in Table 3.2.

DISCUSSION

In surgical education, due to the complex structure of training and evaluation, several studies have explored the reliability of self-assessment using various methodologies [5, 10, 18]. In the past decade, VR simulators have gained significance in surgical skills

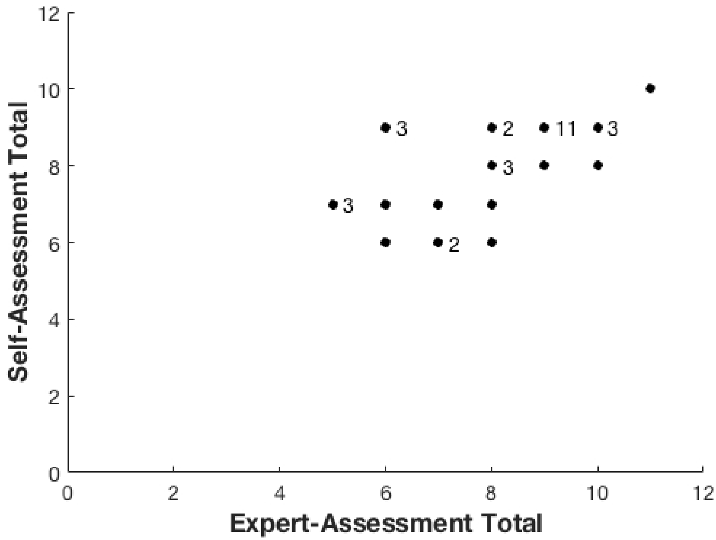


Figure 3.2: Self-assessment (SA) versus expert-assessment (EA) score for aggregated responses to 'usage of instruments' questions. Numbers to the right of data points show the number of coincident data points at the same coordinates, i.e., the number of people with the same combination of SA and EA scores

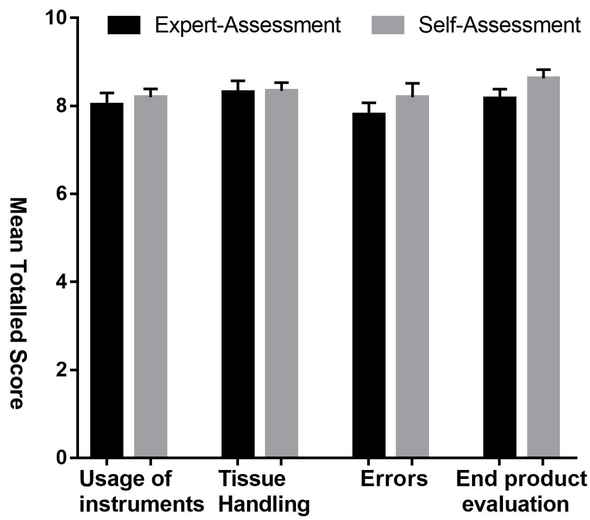


Figure 3.3: Percentage histogram showing the (qualitative similarity of the) overall distribution of responses from expert-assessment (black) and self-assessment (grey)

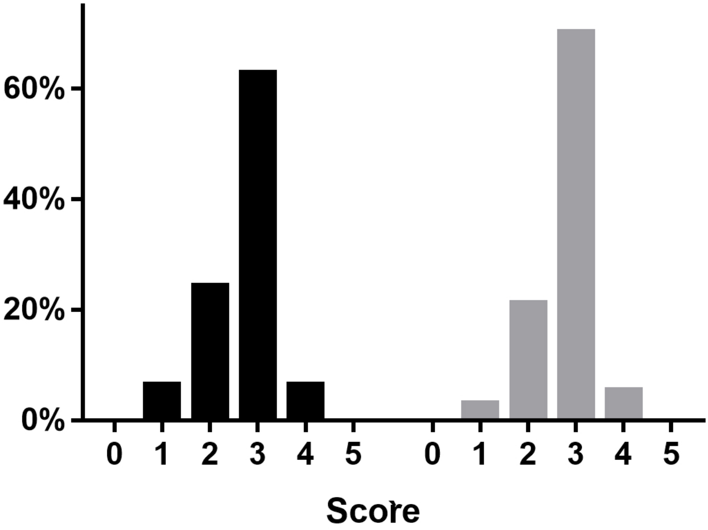


Figure 3.4: Mean ±SEM for the expert-assessment (black) and the self-assessment (grey) total score for the four question groups described on the x-axis

Criteria	Mean of expert-assessment (SD; SEM)	Mean of self-assessment (SD; SEM)	Spearman's rank correlation (p value)	Sum of signed ranks (W) (p value)
All criteria	32.31 (5.05; 0.85)	33.37 (4.31; 0.73)	0.6431 (<0.0001*)	116 (0.1667)
Usage of instruments	8.03 (1.60; 0.27)	8.20 (1.13; 0.19)	0.6208 (<0.0001*)	38 (0.4760)
Tissue handling and usage of NDH	8.31 (1.53; 0.26)	8.34 (1.1; 0.19)	0.3042 (0.0378*)	35 (0.9753)
Errors	7.80 (1.62; 0.27)	8.20 (1.86; 0.31)	0.5376 (0.0004*)	87 (0.1888)
End-product evaluation	8.17 (1.27; 0.21)	8.62 (1.17; 0.20)	0.5180 (0.0007*)	119 (0.0339*)

Table 3.1: Statistics comparing overall and grouped self-assessment with expert-assessment

	Eindhoven, The Netherlands	Cluj-Napoca, Romania	Rajahmundry, India	Total
Male	4	3	11	18
Female	5	2	10	17
Total	9	5	21	35

Table 3.2: Demographic data of participants

training and assessment; and a number of studies prove that they provide feedback that is quite essential for the participants to self-assess their performance [19, 20]. For surgical specialties self-assessment to be more accurate, Mandel et al. [21] suggest that the use of task specific and global check lists should be incorporated. Moreover, as Kostons et al. [7] mentioned in their review on self-assessment, when concurrent monitoring is hampered, that is likely over a period of time, learners have poor recollection of their performance which in turn may hamper their self-assessment after the task.

The objective of this study was to encompass the findings of these prominent studies in surgical training and incorporate them into the study design. Whilst these studies have established the importance of self-assessment as a methodology and its role in education and training, this is the first which has focussed on evaluating performance in individual components of the task. Therefore, the surgical residents were trained on VR simulators, self-assessment was done immediately after the procedure using the CAT form, and they participated in a curriculum that detailed the procedural tasks of the LC.

Evaluating the responses to all components taken together agreed with previous studies: there is a strong correlation between the aggregated responses to the evaluation given by the participants and experts. Evaluating individual procedural tasks independently allowed for individual insights on the strengths and weaknesses in performance and evaluation. The fact that the results indicated a strong correlation between expert- and self-assessment in terms of the 'use of instruments' category could be attributed to the training on VR simulators and box trainers prior to the procedure. A strong correlation found in the evaluation of 'errors' category might indicate a clear layout of errors in the CAT form. Evaluating the distribution of differences leads to no significant differences between the means of the distribution except in the case of the end-product evaluation.

The weaker correlation in terms of tissue handling and usage of NDH could probably be explained by difficulties in observing the NDH, as most surgeons are inclined to look at the actions performed with their dominant hand. The significant difference in the difference of means in the 'end point evaluation' may be attributed to lack of adequate focus on these aspects during the curriculum. Overall, however, the distribution of self-assessment scores is similar and well correlated with expert-assessment. This suggests that self-assessment is a reliable tool to assess one's own performance.

The limitation of our study was the lack of consistent instruction on the usage of the CAT tool to the participants prior to self-assessment. A few studies suggest that surgical residents are better able to self-assess their performance after they have watched benchmark videos; moreover, courses concentrated on the procedural skills of the task have been shown to significantly improve the outcomes of the self-assessment of surgical residents [22, 23].

We intend to explore further how self-assessment is integrated into surgical curricula and, in particular, to investigate whether providing videos and/or images as reference for those conducting self-assessment could improve the efficacy of self-assessment in the areas we found to be less matched with expert-assessment. This in turn could prove beneficial in providing more accurate formative and summative self-assessment in laparoscopic surgical skills.

CONCLUSION

Provided that there is proper understanding and training of the evaluation criteria beforehand, self-assessment using the CAT form gives results that are consistently not different from expert-assessment when assessing one's proficiency in surgical skills. Areas where there was less agreement could be explained by variations in training.

REFERENCES

[1] K. W. Eva and G. Regehr, *"I'll never play professional football" and other fallacies of self-assessment,* Journal of Continuing Education in the Health Professions **28**, 14 (2008).

[2] D. Boud, *Avoiding the traps: seeking good practice in the use of self assessment and reflection in professional courses,* Social Work Education **18**, 121 (1999).

[3] D. Boud, *Enhancing Learning Through Self-Assessment* (Routledge, 1995).

[4] D. Boud and N. Falchikov, *Quantitative studies of student self-assessment in higher education: a critical analysis of findings,* Higher Education **18**, 529 (1989).

[5] R. M. Epstein, *Assessment in medical education,* New England Journal of Medicine **356**, 387 (2007).

[6] M. Ward, L. Gruppen, and G. Regehr, Advances in Health Sciences Education **7**, 63 (2002).

[7] D. Kostons, T. van Gog, and F. Paas, *Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning,* Learning and Instruction **22**, 121 (2012).

[8] F. Dochy, M. Segers, and D. Sluijsmans, *The use of self-, peer and co-assessment in higher education: A review,* Studies in Higher Education **24**, 331 (1999).

[9] K.-P. Kwan and R. W. Leung, *Tutor versus peer group assessment of student performance in a simulation training exercise,* Assessment & Evaluation in Higher Education **21**, 205 (1996).

[10] M. J. Gordon, *Self-assessment programs and their implications for health professions training,* Academic Medicine **67**, 672 (1992).

[11] D. A. Davis, P. E. Mazmanian, M. Fordis, R. V. Harrison, K. E. Thorpe, and L. Perrier, *Accuracy of physician self-assessment compared with observed measures of competence,* JAMA **296**, 1094 (2006).

[12] R. Aggarwal, K. Moorthy, and A. Darzi, *Laparoscopic skills training and assessment,* British Journal of Surgery **91**, 1549 (2004).

[13] J. R. Eriksen and T. Grantcharov, *Objective assessment of laparoscopic skills using a virtual reality stimulator,* Surgical Endoscopy **19**, 1216 (2005).

[14] T. P. Grantcharov, V. B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen, *Randomized clinical trial of virtual reality simulation for laparoscopic skills training,* British Journal of Surgery **91**, 146 (2004).

[15] M. Nagendran, K. S. Gurusamy, R. Aggarwal, M. Loizidou, and B. R. Davidson, *Virtual reality training for surgical trainees in laparoscopic surgery,* Cochrane Database of Systematic Reviews (2013), 10.1002/14651858.cd006575.pub3.

[16] N. Francis, A. Fingerhut, R. Bergamaschi, and R. Motson, eds., *Training in Minimal Access Surgery* (Springer London, 2015).

[17] D. Miskovic, M. Ni, S. M. Wyles, R. H. Kennedy, N. K. Francis, A. Parvaiz, C. Cunningham, T. A. Rockall, A. M. Gudgeon, M. G. Coleman, and G. B. Hanna, *Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England,* Annals of Surgery **257**, 476 (2013).

[18] S. Arora, D. Miskovic, L. Hull, K. Moorthy, R. Aggarwal, H. Johannsson, S. Gautama, R. Kneebone, and N. Sevdalis, *Self vs expert assessment of technical and non-technical skills in high fidelity simulation,* The American Journal of Surgery **202**, 500 (2011).

[19] J. MacDonald, R. G. Williams, and D. A. Rogers, *Self-assessment in simulation-based surgical skills training,* The American Journal of Surgery **185**, 319 (2003).

[20] D. A. Rogers, G. Regehr, T. R. Howdieshell, K. A. Yeh, and E. Palm, *The impact of external feedback on computer-assisted learning for surgical technical skill training,* The American Journal of Surgery **179**, 341 (2000).

[21] L. S. Mandel, B. A. Goff, and G. M. Lentz, *Self-assessment of resident surgical skills: Is it feasible?* American Journal of Obstetrics and Gynecology **193**, 1817 (2005).

[22] N. Tanoue, L. N. Korovin, M. Carton, C. A. Galvani, and I. Ghaderi, *Faculty feedback versus residents' self-assessment of operative performance: Different but complementary,* The American Journal of Surgery **215**, 288 (2018).

[23] R. A. Stewart, L. S. Hauge, R. D. Stewart, R. L. Rosen, A. Charnot-Katsikas, and R. A. Prinz, *A CRASH course in procedural skills improves medical students' self-assessment of proficiency, confidence, and anxiety,* The American Journal of Surgery **193**, 771 (2007).

4

"REFLECTION-BEFORE-PRACTICE" IMPROVES SELF-ASSESSMENT AND END-PERFORMANCE IN LAPAROSCOPIC SURGICAL SKILLS TRAINING

S.Ganni, S.M.B.I. Botden, D.P.Schaap, B.H.Verhoeven, R.H.M.Goossens, J.J.Jakimowicz

Published in Journal of Surgical Education, 75(2), 527-533, August 2017

4

Abstract

Objective

To establish whether a systematized approach to self-assessment in a laparoscopic surgical skills course improves accordance between expert- and self-assessment.

Design

A systematic training course in self-assessment using Competency Assessment Tool was introduced into the normal course of evaluation within a Laparoscopic Surgical Skills training course for the test group ($n = 30$). Differences between these and a control group ($n = 30$) who did not receive the additional training were assessed.

Setting

Catharina Hospital, Eindhoven, The Netherlands ($n = 27$), and GSL Medical College, Rajahmundry, India ($n = 33$).

Participants

Sixty postgraduate year 2 and 3 surgical residents who attended the 2-day Laparoscopic Surgical Skills grade 1 level 1 curriculum were invited to participate.

Results

The test group ($n = 30$) showed better accordance between expert- and self-assessment (difference of 1.5, standard deviation [SD] = 0.2 versus 3.83, SD = 0.6, $p = 0.009$) as well as half the number (7 versus 14) of cases of over-reporting. Furthermore, the test group also showed higher overall mean performance (mean = 38.1, SD = 0.7 versus mean = 31.8, SD = 1.0, $p < 0.001$) than the control group ($n = 30$). The systematic approach to self-assessment can be viewed as responsible for this and can be seen as "reflection-before-practice" within the framework of reflective practice as defined by Donald Schon.

Conclusion

Our results suggest that "reflection-before-practice" in implementing self-assessment is an important step in the development of surgical skills, yielding both better understanding of one's strengths and weaknesses and also improving overall performance.

Introduction

The development of technical skills is crucial for surgical residents and surgeons. Simulation-based training is a very important tool to enhance this competence. Besides supervised teaching by expert, trained surgeons, self-assessment and self-directed learning are key elements in surgical training.[1, 2] Several studies have shown that integration of self-assessment is beneficial for the development of a surgeon's career.[3, 4]

There is disagreement in terms of the desirable role of self-assessment, between the literature on self-assessment theories and that concerning its real-world implementation in surgical practice. The theoretical literature tends to focus on the use of self-assessment as a means of improving reflective practice and thereby improving the individual's overall professional competence and skills.[5, 6] Evaluation of real-world self-assessment in surgical practice often focuses on trying to achieve accordance between expert and self-assessment and a reduction in overestimation of performance.[7]

Although self-assessment has been considered a vital component for professional self-regulation and development for a long time, many studies debate the effectiveness and efficacy of self-assessment in skills training and state that there is room for improvement.[5, 8–10] Recently, several authors, such as Ward et al[11] propose that resolving weaknesses in the methodologies used to evaluate self-assessment would yield a more positive evaluation of self-assessment's efficacy. Because of these improved methodologies, it has been shown that trainees or surgical residents are in fact able to self-assess their weaknesses and strengths similarly to expert assessment.[7, 12, 13]

Regardless of the field using self-assessment, the ideal is to improve the ability of individual candidates to accurately assess their own ability with the aim to improve their overall performance; to this end, many tools and methodologies have been suggested for the improvement of self-assessment itself.[6, 14, 15] One of the most important conclusions is that surgical residents assess their own procedural performance more accurately after watching benchmark videos of expert performances and their own performances.[8, 16] Stewart et al indicated a concentrated, intense course in procedural skills before evaluation for self-assessment to be more accurate, namely greater accordance between expert- and self-assessment.[17]

This study aimed to determine whether implementing a self-assessment training tool in a validated laparoscopic surgical skills course will improve the accordance between self- and expert assessment.

Materials and Methods

Participants

Sixty surgical residents who attended the 2-day Laparoscopic Surgical Skills (LSS) grade 1 level 1 curriculum were invited to participate in 2 centers: Catharina Hospital, Eindhoven, The Netherlands (n = 27), and GSL Medical College, Rajahmundry, India (n = 33). Their expertise level ranged from postgraduate year 2 to 3. All participants voluntarily enrolled in the study and signed an informed consent before the start of the curriculum. All participants had completed and passed an online examination on the basics of laparoscopic surgery to be eligible for participation in the program. Each participant completed a questionnaire with questions pertaining to demographics, experience in laparoscopic surgery, and time spent preparing for the curriculum.

Assessment tool

The Competency Assessment Tool (CAT) used in this study is an operation-specific assessment tool that was adapted and validated for the laparoscopic cholecystectomy (LC) procedure for use within the LSS curriculum.[18] In this study, it was used as a tool for self- and expert assessment. The CAT evaluation criteria are spread across 3 procedural tasks:

1. exposure of both the cystic artery and cystic duct,

2. cystic pedicle dissection, and

3. resection of the gallbladder from the liver bed.

Within these tasks, performance is rated on a five-point, task-specific scale based on the efficient usage of instruments, the handling of tissue with the non-dominant hand, errors within each task, and the end-product of each task. A maximum of 48 points can be scored on the CAT assessment, and a total score of 30 or more was considered a pass for the LC course. Four expert surgeons, 2 from each of the respective locations conducting the curriculum, were invited to participate as expert assessors for both the test and control groups. They all had previous experience in using the CAT form for evaluation. Their laparoscopic surgical experience ranged from 5 to 25 years, each with more than 200 laparoscopic procedures performed as main surgeon. The surgeons were not aware whether the candidates they were assessing had the additional training or not when conducting their assessment.

PROTOCOL

All participants completed the standard training and instructions of the LSS grade 1 level 1 curriculum. During the course, they received an interactive discursive training with experts on the basics of laparoscopic surgery, LC, virtual reality simulators, and box trainers. The participants were divided in 2 groups based on the days they attended courses; into a test group ($n = 30$) and a control group ($n = 30$). The participants of both groups were instructed by the expert surgeons on the procedural tasks of the LC. Immediately before they performed the procedure, the test group received an additional training session on self-assessment (Figure 4.1).

This session totalled 30 minutes in duration and started with the instructor introducing the theoretical meaning and professional benefits observed in the literature of self-assessment. The group was then given the CAT form and instructed to read it. Each criterion was explained in detail by the instructor. The relation between the word-based definitions on the CAT form and their score-based equivalents was explained. The instructor then held a question and answer session to resolve any of the participant's concerns. Where possible, the criteria were accompanied. by illustrative videos, showing examples of both good (CAT score of 4) and bad (CAT score of 1) practice, for additional explanation and images of the same were printed overleaf the CAT form for later reference (Figure 4.2). As is the current norm, the control group were given the CAT form just before the procedure.

Thereafter, every participant performed the procedural tasks of the LC on a porcine

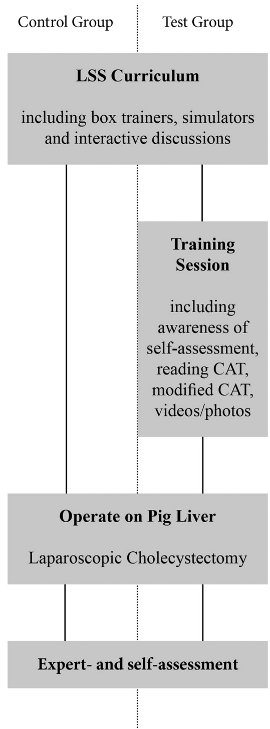


Figure 4.1: An illustration of the training protocol depicting the differences between the test and control groups.

liver placed in a box trainer. The box trainer was placed on a height-adjustable table with an ergonomically correct position of monitors and instruments. The entry ports for the laparoscopic instruments mimicked the incision points in the clinical setting. A fellow participant played the role of surgical assistant during the procedure. The assessors were asked not to provide feedback on the participant’s performance during the procedure. Immediately after the procedure, each participant and an expert observer completed a CAT form independently of each other.

STATISTICAL ANALYSIS

The analysis was performed by comparing the differences between the expert- and self-assessment scores between the test and control groups based on the aforementioned criteria within the procedural tasks. Statistical and absolute differences were calculated between expert- and self-assessment scores using MATLAB (R16b), and the obtained data were analysed and presented using GraphPad Prism (Version 7.00). Because the data were non-parametric, the Mann-Whitney U-test was used to calculate significant

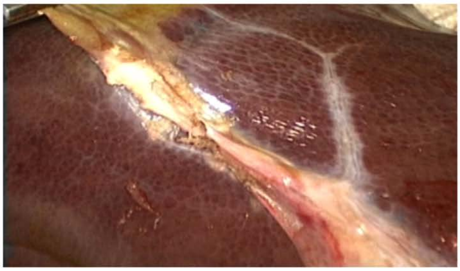
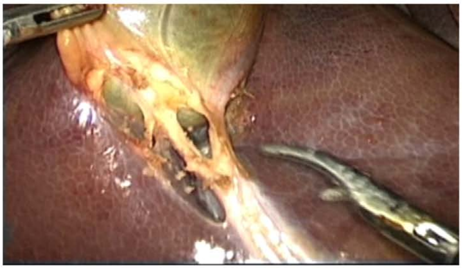
End-product: Exposure of cystic duct and artery	
	
The anatomical structures are not identified and thus corresponds to a CAT Score of 1.	Here, there is crystal clear demonstration of the anatomy which corresponds to a CAT Score of 4.

Figure 4.2: An example of the photographic reminders of the CAT scores corresponding to both good and bad practices which were overleaf the modified CAT form given to the test group participants.

differences between the assessment scores. Other statistical differences were calculated using Graphpad Prism. A $p < 0.05$ was considered statistically significant.

Both the numerical difference and the absolute difference between expert assessment and self-assessment scores were used. When only using one of these measures, clinically relevant correlations or differences could get lost. For example, if part of the trainees score themselves higher and the other half lower than the experts, this is clinically relevant, but the mean of the self-assessment would be equal to the expert assessment. An improvement in the quality of self-assessment can be seen if the numerical difference between self-assessment and expert assessment is closer to zero, when comparing the test group to the control group. If the absolute difference is smaller, this corresponds to an improvement in self-assessment after self-assessment training.

RESULTS

ASSESSMENT SCORE

The total overall score given by the experts for the test group in the CAT assessment shows a significantly higher mean than for the control group (mean = 38.1, standard deviation [SD] = 0.7 versus mean = 31.8, SD = 1.0, $p < 0.001$) (Figure 4.3). In the control group, 9 participants scored below 30 (regarded as a fail for the course) on the expert assessment, whereas no participant scored less than 30 in the test group. The same

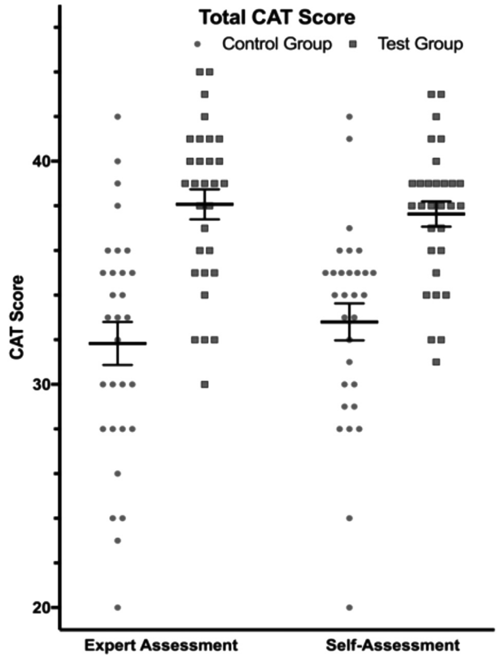


Figure 4.3: The total score on the CAT form for each participant as assessed by an expert.

pattern is seen in the self-assessment results, with a mean of 37.6 (SD = 0.6) for the test group and mean of 32.8 (SD = 0.8) for the controls ($p < 0.0001$). On self-assessment, 7 participants of the control group scored less than 30, compared to none in the test group. The scores of the individual tasks, for both expert- and self-assessment (Table 1), without significant differences in outcome between the test and control group on the separate scored items for both expert- and self-assessment. However, when it comes to the use of tools in the cystic pedicle dissection, the self-assessment was 0.3 points higher on average for the control group compared with the expert assessment, while their expert-assessed performances were the lowest of all (Table 4.1).

Figure 4.4 shows the distribution of the differences between the total expert-assessed and the self-assessed CAT scores. The interquartile range and SD between the scores are much smaller in the test group than in the control group. Also, the mean absolute difference between expert- and self-assessment is significantly lower in the test group compared to the control group (1.5 versus 3.83), with a smaller SD (0.2 versus 0.6) (p = 0.009) (Table 4.2). In addition, the number of overestimated performances decreased from 14 in the control group to 7 in the test group (Figure 4.4).

Looking at the absolute difference between expert- and self-assessment (Figure 4.5)

	Total EA	Total SA	Test EA	Test SA	Control EA	Control SA
Use of graspers and tools						
Exposure of cystic duct and artery	2.8 (0.58)	2.7 (0.45)	3.0 (0.37)	2.9 (0.25)	2.6 (0.67)	2.6 (0.56)
Cystic pedicle dissection	2.8 (0.69)	2.8 (0.52)	3.2 (0.55)	3.0 (0.56)	2.4 (0.63)	2.7 (0.45)
Resection of gallbladder	3.0 (0.74)	2.8 (0.64)	3.2 (0.66)	3.1 (0.59)	2.8 (0.76)	2.6 (0.56)
Tissue handling						
Exposure of cystic duct and artery	2.9 (0.59)	2.8 (0.56)	3.2 (0.43)	3.1 (0.35)	2.6 (0.61)	2.6 (0.61)
Cystic pedicle dissection	2.9 (0.70)	2.9 (0.58)	3.2 (0.68)	3.0 (0.61)	2.7 (0.64)	2.9 (0.55)
Resection of gallbladder	3.0 (0.65)	3.0 (0.64)	3.3 (0.53)	3.3 (0.48)	2.8 (0.66)	2.6 (0.61)
Errors						
Exposure of cystic duct and artery	2.9 (0.64)	2.9 (0.69)	3.2 (0.50)	3.1 (0.53)	2.7 (0.65)	2.7 (0.75)
Cystic pedicle dissection	2.8 (0.71)	2.9 (0.75)	3.1 (0.63)	3.1 (0.65)	2.5 (0.68)	2.8 (0.81)
Resection of gallbladder	2.6 (1.05)	2.9 (0.86)	2.9 (0.87)	3.1 (0.70)	2.4 (1.17)	2.6 (0.93)
End Evaluation						
Exposure of cystic duct and artery	2.9 (0.56)	3.0 (0.58)	3.1 (0.46)	3.2 (0.50)	2.6 (0.55)	2.7 (0.57)
Cystic pedicle dissection	2.9 (0.70)	2.9 (0.54)	3.2 (0.68)	3 .0 (0.61)	2.6 (0.61)	2.9 (0.45)
Resection of gallbladder	2.9 (0.80)	3.0 (0.65)	3.2 (0.81)	3.2 (0.25)	2.7 (0.74)	2.8 (0.55)

Table 4.1: Expert-Assessment (EA) and Self-Assessment (SA) Scores on the Separate Scored Aspects: The Mean (standard deviation) Response for Each Criterion on the CAT Form is Shown for Both the Total Group and Test/Control Group Only. Differences Between Expert- and Self-Assessment in Total CAT Scores

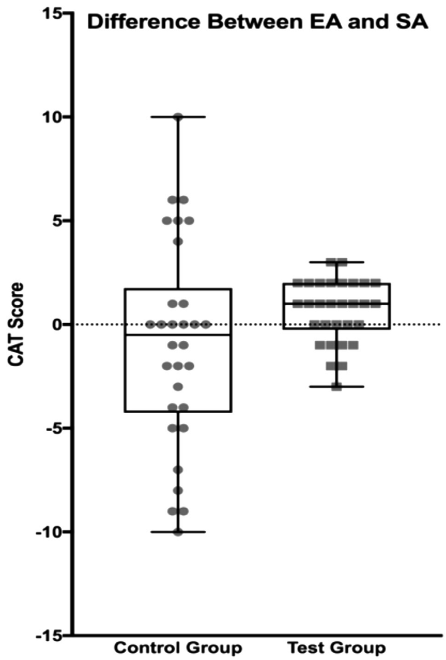


Figure 4.4: Box-plot of the difference between expert- and self- assessment scores, by means of deviation of the individual assessment scores between the two groups ($p = 0.045$).

	Difference between EA vs SA			**Absolute difference EA vs SA**		
	Test group	*Control group*	*p-value*	*Test group*	*Control group*	*p-value*
All criteria	0.63 (1.52; 0.28)	-0.97 (4.97; 0.91)	0.045*	1.5 (1.11; 0.20)	3.83 (3.24; 0.59)	0.009*
Usage of instruments	0.27 (0.91; 0.17)	-0.10 (1.37; 0.25)	0.207	0.80 (0.48; 0.09)	0.97 (0.96; 0.18)	0.841
Tissue handling and usage of NDH	0.60 (0.50; 0.09)	0.00 (1.74; 0.32)	0.040*	0.6 (0.50; 0.09)	1.27 (1.17; 0.21)	0.019*
Errors	-0.20 (0.81; 0.15)	-0.43 (1.94; 0.35)	0.561	0.6 (0.56; 0.10)	1.43 (1.36; 0.25)	0.011*
End-product evaluation	0.10 (0.80; 0.15)	-0.43 (1.33; 0.24)	0.047*	0.63 (0.49; 0.09)	1.03 (0.93; 0.17)	0.097

Table 4.2: Summative statistics (mean [standard deviation; standard-error in the mean]) for the numeric and absolute difference between the expert assessment (EA) and self-assessment (SA) between the test and the control group. Significant differences are calculated using the Mann-Whitney U-test, with $p < 0.05$ considered a significant difference.

4

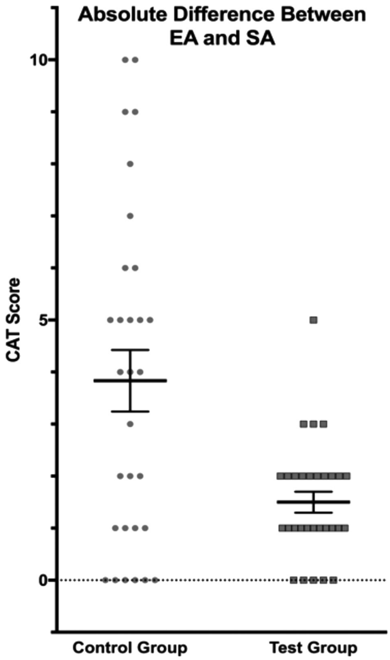


Figure 4.5: The absolute value of the difference between expert-and self-assessment. Individual difference scores are shown as gray points, with the mean ± SEM shown in black.

shows that the mean is much lower in the test group -1.5 ± 0.2 versus 3.83 ± 0.6 ($p = 0.009$). Table 4.2 explains the differences between these values for the grouped items (usage of instruments, tissue handling, errors, and end-product evaluation). Significant differences are seen in all items, except for the usage of instruments, and the items on this subject were scored equal in both groups. As also shown in the table, the calculated difference between self- and expert assessment of tissue handling was 0.0 in the control group; however, the absolute differences shows 1.27.

Discussion

It is important that doctors and, in particular, surgeons know how they perform during surgical procedures. If a surgeon is not aware of possible hazardous movements or near-incidents, this could result in unnecessary high complication rates. Therefore, it is important that surgeons are assessed on their skills before they perform unsupervised procedures in the clinical setting. When surgeons are accurately aware of their own skills level, with strengths. and weaknesses, they know what skills are important to practice more extensively. Therefore, self-assessment could be an important step in the development of surgical skills and enhance patient safety.

This study aimed to assess whether implementing a self-assessment training tool, which includes the latest methodology recommendations, improves the accordance between self- and expert assessment in a validated laparoscopic surgical skills course. This resulted in a single training session immediately before performing the procedure to allow the candidates time to reflect on the assessment form and its criteria based on which their skills were assessed. Additionally, it drew attention to the fact that effective self- assessment was for their long-term professional benefit. Furthermore, the recommendations of previously described literature were implemented through an additional training session by means of videos and photographs of both good and bad practice.[6, 12–15]

While it is possible to expect that training in how to use the CAT form would yield the better accordance between expert- and self-assessment seen in the results, it is interesting to note that this systematic approach to implementing self-assessment also improved the candidates' overall performance. Another benefit was a reduction in the number of candidates in the test group who over-reported their performance, suggesting that the training session may have made the participants more aware of their proficiency.

The candidates in the test group may also have benefited from considering the exact assessment criteria before assessing themselves, resulting in a better understanding of both the criteria themselves and the relationship between their word-based definition and their score-based equivalents. Moreover, the value of both educating and motivating the candidates with the short- and long-term benefits of accurately applying self-assessment to their practice resulted in increased performance outcomes across almost all evaluation criteria.

The concept of reflective practice, as it is currently understood, beholds 3 categories defined by Donald Schon: knowing-in-action, reflection-on-action, and reflection-in-action.[19] However, in implementing this improved approach to self-assessment, it seems we have a form of "reflection-before-practice" here. That is, providing the candidates with an understanding of the individual professional benefits of self-assessment, as well as examples of both good and bad practice, during and before they self-assess. Furthermore, providing the candidates with a clear objective of the expected outcomes in advance could have created a constant reflection of these objectives during the course of their performance. It is this that appears to have resulted in improved overall performance and increased accordance between expert- and self-assessment.

The scope of this study is limited by its implementation on a relatively basic laparoscopic procedure and assessment. More comprehensive assessment tools such as the observational clinical human reliability assessment (OCHRA) have been highly regarded in assessing not only the competency of skills but also human reliability by means of consequential and non-consequential errors during a surgical procedure.[20] Further research should be done to investigate whether "reflection-before-practice" using assessment tools such as OCHRA in laparoscopic procedures would also improve both competency and human reliability factor of surgeons. Furthermore, establishing whether similar improvements are seen when this approach is applied to a variety of surgical procedures would prove the approach to be an effective way of improving both self-assessment and, importantly, performance.

CONCLUSION

As might have been expected, candidates who receive training in self-assessment in surgical skills training significantly improve their accordance between self- and expert assessment and a reduction in over-reporting. Here, however, a second function of the training in self-assessment is seen; the participants improved their overall performance.

Thus, training in self-assessment, seen as "reflection-before-practice," can be used to improve not only the accuracy of self-assessment but also the actual surgical performance.

4

REFERENCES

[1] D. Kostons, T. van Gog, and F. Paas, *Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning,* Learning and Instruction **22**, 121 (2012).

[2] E. M. Kwasnik and J. Carter, *To see ourselves as others see us: self-assessment in surgical residency,* Current Surgery **56**, 145 (1999).

[3] S. Arora, D. Miskovic, L. Hull, K. Moorthy, R. Aggarwal, H. Johannsson, S. Gautama, R. Kneebone, and N. Sevdalis, *Self vs expert assessment of technical and non-technical skills in high fidelity simulation,* The American Journal of Surgery **202**, 500 (2011).

[4] M. J. Gordon, *Self-assessment programs and their implications for health professions training,* Academic Medicine **67**, 672 (1992).

[5] D. Boud, *Avoiding the traps: seeking good practice in the use of self assessment and reflection in professional courses,* Social Work Education **18**, 121 (1999).

[6] D. Boud, *Enhancing Learning Through Self-Assessment* (Routledge, 1995).

[7] S. Ganni, M. K. Chmarra, R. H. M. Goossens, and J. J. Jakimowicz, *Self-assessment in laparoscopic surgical skills training: Is it reliable?* Surgical Endoscopy **31**, 2451 (2016).

[8] M. J. Gordon, *A review of the validity and accuracy of self-assessments in health professions training,* Academic Medicine **66**, 762 (1991).

[9] L. Arnold, T. L. Willoughby, and E. V. Calkins, *Self-evaluation in undergraduate medical education,* Academic Medicine **60**, 21 (1985).

[10] K. W. Eva, J. P. Cunnington, H. I. Reiter, D. R. Keane, and G. R. Norman, *How can I know what I don't know? Poor self assessment in a well-defined domain,* Advances in Health Sciences Education **9**, 211 (2004).

[11] M. Ward, H. MacRae, C. Schlachta, J. Mamazza, E. Poulin, R. Reznick, and G. Regehr, *Resident self-assessment of operative performance,* The American Journal of Surgery **185**, 521 (2003).

[12] J. MacDonald, R. G. Williams, and D. A. Rogers, *Self-assessment in simulation-based surgical skills training,* The American Journal of Surgery **185**, 319 (2003).

[13] L. S. Mandel, B. A. Goff, and G. M. Lentz, *Self-assessment of resident surgical skills: Is it feasible?* American Journal of Obstetrics and Gynecology **193**, 1817 (2005).

[14] R. Jamshidi, T. LaMasters, D. Eisenberg, Q.-Y. Duh, and M. Curet, *Video self-assessment augments development of videoscopic suturing skill,* Journal of the American College of Surgeons **209**, 622 (2009).

[15] S. K. Sarker, R. Hutchinson, A. Chang, C. Vincent, and A. W. Darzi, *Self-appraisal hierarchical task analysis of laparoscopic surgery performed by expert surgeons,* Surgical Endoscopy **20**, 636 (2006).

[16] Y. Hu, D. Tiemann, and L. M. Brunt, *Video self-assessment of basic suturing and knot tying skills by novice trainees,* Journal of Surgical Education **70**, 279 (2013).

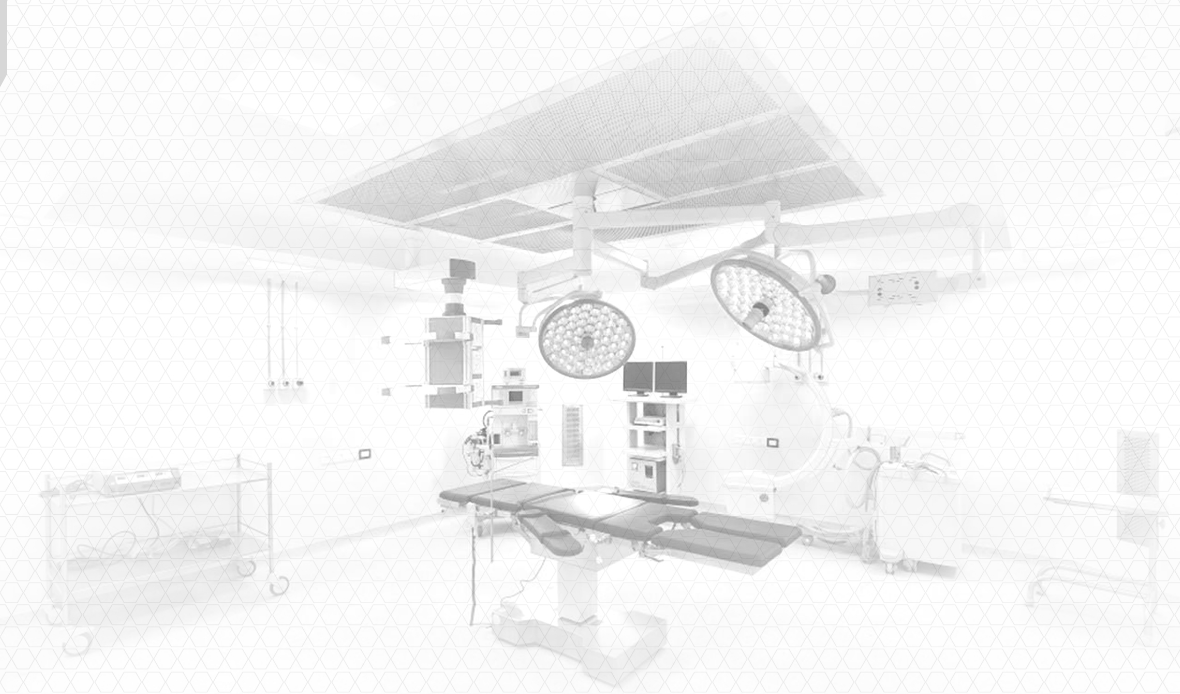
[17] R. A. Stewart, L. S. Hauge, R. D. Stewart, R. L. Rosen, A. Charnot-Katsikas, and R. A. Prinz, *A CRASH course in procedural skills improves medical students' self-assessment of proficiency, confidence, and anxiety,* The American Journal of Surgery **193**, 771 (2007).

[18] D. Miskovic, M. Ni, S. M. Wyles, R. H. Kennedy, N. K. Francis, A. Parvaiz, C. Cunningham, T. A. Rockall, A. M. Gudgeon, M. G. Coleman, and G. B. Hanna, *Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England,* Annals of Surgery **257**, 476 (2013).

[19] D. A. Schön, *The Reflective Practitioner* (Routledge, 2017).

[20] B. Tang, *Identification and categorization of technical errors by observational clinical human reliability assessment (OCHRA) during laparoscopic cholecystectomy,* Archives of Surgery **139**, 1215 (2004).

5



COMPETENCY ASSESSMENT TOOL FORLAPAROSCOPIC SUTURING: DEVELOPMENT AND RELIABILITY EVALUATION

W.M.Ijgosse, E. Leijte, S.Ganni, J.Luursema, N.K.Francis, J.J.Jakimowicz, S.M.B.I.Botden

Published in Surgical Endoscopy, August 2019

5

Abstract

Background Laparoscopic suturing can be technically challenging and requires extensive training to achieve competency. To date no specific and objective assessment method for laparoscopic suturing and knot tying is available that can guide training and monitor performance in these complex surgical skills. In this study we aimed to develop a laparoscopic suturing competency assessment tool (LS-CAT) and assess its inter-observer reliability.

Methods We developed a bespoke CAT tool for laparoscopic suturing through a structured, mixed methodology approach, overseen by a steering committee with experience in developing surgical assessment tools. A wide Delphi consultation with over twelve experts in laparoscopic surgery guided the development stages of the tool. Following, subjects with different levels of laparoscopic expertise were included to evaluate this tool, using a simulated laparoscopic suturing task which involved placing of two surgical knots. A research assistant video recorded and anonymised each performance. Two blinded expert surgeons assessed the anonymised videos using the developed LS-CAT. The LS-CAT scores of the two experts were compared to assess the inter-observer reliability. Lastly, we compared the subjects' LS-CAT performance scores at the beginning and end of their learning curve.

Results This study evaluated a novel LS-CAT performance tool, comprising of four tasks. Thirty-six complete videos were analysed and evaluated with the LS-CAT, of which the scores demonstrated excellent inter-observer reliability. Cohen's Kappa analysis revealed good to excellent levels of agreement for almost all tasks of both instrument handling and tissue handling (0.87; 0.77; 0.75; 0.86; 0.85, all with $p < 0.001$). Subjects performed significantly better at the end of their learning curve compared to their first attempt for all LS-CAT items (all with $p < 0.001$).

Conclusions We developed the LS-CAT, which is a laparoscopic suturing grading matrix, with excellent inter-rater reliability and to discriminate between experience levels. This LS-CAT has a potential for wider use to objectively assess laparoscopic suturing skills.

INTRODUCTION

Over the past two decades, Minimal Invasive Surgery (MIS) has expanded rapidly with more advanced surgical operations now being performed laparoscopically. This often involves carrying out reconstructive procedures which requires the skills of performing laparoscopic suturing [1, 2].

Training for laparoscopic suturing is an integral part of the laparoscopic surgical curriculum [3] and has moved from the operating room to a skills lab setting [4]. Complex surgical skills such as laparoscopic suturing and knot tying are challenging due to the inherent limitations of MIS such as an altered depth perception, two-dimensional vision, ergonomic issues and the small working field [5, 6].

Extensive training, therefore, is required to overcome these limitations and to achieve competency and is often based on the principle of modelling, repetitive practice and formative feedback [7]. Surgical residents are currently more and more restricted in their clinical working hours, reducing their opportunities for gaining practical surgical experience. Therefore, assessment of performance is required not only to ensure competency but to guide and enhance the efficiency of learning [8]. Assessment of laparoscopic suturing is traditionally dependent on subjective evaluation by trainers since objective evaluation has not yet been established.

Several attempts to objectively assess laparoscopic suturing have been reported in literature including the use of virtual reality (VR) simulation, motion-tracking systems or check lists. The application of VR to objectively evaluate laparoscopic suturing skills can be challenging [3]. VR simulators are able to fully assess the trainees, but lack the important haptic feedback, needed for laparoscopic suturing [8]. There are several studies which applied a motion-tracking system to real-time performance [6, 9] to objectively appraise the operative performance of this complex task, but this method is of limited generalisability and external validity. There are various other measurement tools avail able, but they vary in their objectivity, validity and reliability [10]. Mandel et al. mentioned the importance of immediate and specific feedback during training and suggests the use of task-specific and global checklists for both learning and self-assessment [11].

A competency assessment tool (CAT) is a method to assess laparoscopic performance, by describing specific steps in the process of the specific task and evaluates both the process of performance (instrument use, tissue handling and committed errors) and the quality of the end product. The CAT tool has been successfully applied to approve the quality of training in the English National Training Programme for laparoscopic colorectal surgery [12]. Considering the importance of laparoscopic suturing and its wide application within the practice of MIS, there is a clear need for an objective assessment tool that can reliably appraise the operative performance of such complex technique. We therefore aimed to develop a bespoke CAT for laparoscopic suturing and assess the reliability of the tool by assessing the inter-observer reliability.

MATERIALS AND METHODS

DEVELOPMENT OF COMPETENCY ASSESSMENT TOOL

The development of the laparoscopic suturing CAT (LS CAT) was performed with a structured, mixed methodology approach and overseen by a steering committee with experience in developing surgical assessment tools and objective assessment of laparoscopic rectal surgery. A wide Delphi consultation with over twelve international experts in laparoscopic suturing guided the development stages of the tool. The steps were standardised and agreed first prior to defining the task areas for assessment with the tool. Based upon an expert consensus, we deconstructed the procedure into a series of constituent steps. The final model of the LS-CAT was adapted from the original CAT for assessing colorectal surgery [12].

Next, we used a semi-structured interview framework allowing the experts freedom to express their thoughts and explore ideas, whilst also enabling the interviewer to ensure the necessary information was covered [13]. Open questions were used to determine what indicators of performance the expert would look for to assess technical performance of laparoscopic suturing. Additionally, for each task area, two video clips were prepared for the expert to reflect upon the technical performance displayed. A research assistant transcribed the interviews verbatim and analysed them using qualitative methods. After coding and grouping of the statements and until thematic saturation was achieved, the thematic analysis was performed. We collated descriptors of poor and proficient performance from the transcripts and triangulated them into the specific procedural tasks to which they applied to generate the assessment metrics for the draft tool.

The draft of the LS-CAT consisted of four agreed task areas, reflecting steps of the procedure described in the expert consensus. Based on the interviews and error analysis, we developed objective descriptors for each task and refined them through discussions amongst the steering group. To describe the quality of technical performance for each domain (four) for each task area (two) a four-point ordinal scale was used, where a lower score indicates a more proficient technical performance and a high score (four) a poor performance. A total LS-CAT score of eight indicates a perfect and proficient performance, because one point was scored on both items in each task, without errors during the performance.

Tool testing

Training setup

Training took place at the Radboud University Medical Center, Nijmegen. During the first training session, a research assistant was available to instruct subjects prior to conducting the laparoscopic suturing tasks. The research assistant video recorded and anonymised each performance but was not involved in the LS-CAT scoring process. Each participant performed the suturing tasks multiple times to train along a learning curve. The LS-CAT was evaluated using the following suturing task:

A standard suturing task. The participant had to place two surgical knots on a suturing pad in a horizontal plane (double wind followed by two single winds to create a secure surgeon's knot) with a standard length of 20-cm thread. If the thread of a suture was too short to reuse after being cut by the research assistant, a new suture would be placed on the suture pad.

Training subjects

Subjects were divided into three groups based on their self-reported laparoscopic experience: (1) novices were subjects without clinical experience but with understanding of the concept of laparoscopy such as medical interns and first-year residents, (2) intermediates with more than ten basic laparoscopic procedures

Figure 5.1: The eoSim-augmented reality laparoscopic simulator interface

performed but less than twenty advanced laparoscopic procedures and (3) experts with more than twenty advanced laparoscopic procedures performed, therefore consisting of residential surgeons in staff. Because the novices were training on their learning curve, the videos of the end of the learning curve were used as a fourth group.

PROTOCOL

All participants signed an informed consent for the video recording of their task performances prior to the start of the training. When all participants finished the training, we analysed 36 videos from the bulk of all participants' performances, after which two blinded expert surgeons completed the LS-CAT independently of each another. Both experts had experience using the original CAT tool [14], but had not used the adapted version for laparoscopic suturing before. Participation was on voluntary basis and subjects received no compensation. No IRB approval was needed for this study.

EQUIPMENT

The eoSim-augmented reality laparoscopic simulator by eoSurgical Ltd., Edinburgh, Scotland, United Kingdom, was used in this study, in a standard setup (Figure 5.1). This setup consisted of the eoSim laparoscopic case with an internal-mounted high-definition camera and standard supplied equipment that consists of laparoscopic instruments, needle holders, a suturing pad, a thread transfer platform and a box with standard exercise equipment, combined with a 15-inch laptop with the required specification as recommended by eoSurgical and the eoSurgical SurgTrac software installed. The tracking camera, that is mounted in the case, was connected to the laptop via USB 2.0 and used to record each performance of the participant. For every participant, the height of the laptop screen was adjusted to the proper height with the laparoscopic box being placed on a standard height table. Participants used a 30-mm curved needle braided thread suture to perform the task.

STATISTICAL ANALYSIS

All statistical analyses were performed with IBM's SPSS statistics v.25 package. First the total scores for instrument handling, tissue handling and the amount of errors were calculated. Following, the inter-observer reliability was assessed by using Cohen's Kappa analysis for the task scores of instrument handling and tissue handling. A $K > 0.75$ was considered as an excellent agreement [15]. The inter-observer reliability for the calculated total scores between the two observers was assessed using the Pearson correlation, on a 2-tailed significance level of $p < 0.01$. An $r > 0.8$ was considered a high correlation [16]. Lastly, the performance scores at the beginning and end of the learning curve within the novice group were compared using the Mann-Whitney U test. This process was conducted by three independent researchers who were not involved in the scoring process using the filled in LS-CAT forms of the observers (EL calculated the total scores, SMBI conducted the statistical analyses, WMIJ repeated both processes as a final check).

RESULTS

DEVELOPMENT OF COMPETENCY ASSESSMENT TOOL

The final LS-CAT is presented in Figure 5.2. Two vertical columns represent task areas, and four horizontal rows represent the performance domains: giving a total of eight separate items which are scored on a scale of 1-4, where a lower score indicates a more proficient technical performance and a total score of eight indicates a perfect and proficient performance. The third column represents the amount of errors which is scored on four domains for each task resulting in 16 separate items.

Four tasks were agreed on and defined from the consensus document for assessment with the tool: (1) pickup needle in correct orientation to make bite; (2) pass needle through two edges of tissue with appropriate bite placement and tissue handling; (3) create first double wind/throw of the knot and tighten correctly and (4) knot tying.

	Observer A				Observer B			
	Mean rank				Mean rank			
	First attempt	Last attempt	*U*	*p*	First attempt	Last attempt	*U*	*p*
Instrument handling	27.06	9.94	8.00	<0.001	27.19	9.81	5.50	<0.001
Tissue handling	26.97	10.03	9.50	<0.001	26.17	10.83	24.00	<0.001
Pickup needle in correct orientation	27.31	9.69	3.50	<0.001	27.00	10.00	9.00	<0.001
Pass needle through edges of tissue	25.64	11.36	33.50	<0.001	25.58	11.42	34.50	<0.001
Create first double throw	25.94	11.06	28.00	<0.001	26.06	10.94	26.00	<0.001
Knot tying	26.17	10.83	24.00	<0.001	25.53	11.47	35.50	<0.001
Total errors	25.94	11.06	28.00	<0.001	25.19	11.81	41.50	<0.001
Total score	27.31	9.69	3.50	<0.001	27.17	9.83	6.00	<0.001

Table 5.1: Scores of the separate items on the LS-CAT. The values are stated in means and standard deviations

	Instruments handling		Tissue handling	
	κ	*p*	κ	*p*
Pickup needle in correct orientation	0.87	<0.001	0.86	<0.001
Pass needle through edges of tissue	0.77	<0.001	0.51	<0.001
Create first double throw	0.73	<0.001	0.85	<0.001
Knot tying	0.75	<0.001	0.73	<0.001

Table 5.2: Inter-rater agreement for the categorical variables calculated with Cohen's Kappa

	Observer A	Observer B	*r*	*p*
Instrument handling	10.0 (2.8)	9.9 (3.0)	0.98	<0.001
Tissue handling	9.4 (2.3)	9.1 (2.4)	0.86	<0.001
Total score	19.4 (4.9)	19.0 (5.2)	0.96	<0.001
Total errors	2.1 (3.2)	1.9 (3.2)	0.99	<0.001
Total assessment score	21.4 (7.1)	20.9 (7.5)	0.98	<0.001

This is calculated with Pearson correlation, on a 2-tailed significance level of $p < 0.01$

Table 5.3: Correlations between the total scores of the items

Objective assessment of errors in suturing and knot tying

Trainee ID ☐☐☐☐☐ Assessor ID ☐☐☐☐☐ Date ☐☐/☐☐/☐☐

☐ Expert Observer ☐ Self Assessment

TASKS	INSTRUMENTS HANDLING (DH)			TISSUE HANDLING (NDH)			OUTCOME	
	USE OF GRASPERS/NEEDLE HOLDERS			GENTLENESS OF TISSUE HANDLING			ERRORS	NUMBER OF ERRORS
PICK UP NEEDLE IN CORRECT ORIENTATION TO MAKE BITE	4	Uncoordinated movements	Stiff and uncontrolled	4	Stagnant	Does not move	Number of dropping the needle	
	3	Hesitant movement	Controlled, but hesitant and ineffective	3	Lagging	Adjusting with delay or without efficiency	Incorrect orientation (needle)	
	2	Skillful movements	Smooth, controlled and meaningful	2	Meaningful	Meaningful adjustment of to improve exposure	Incorrect position (needle/ needle holder)	
	1	Versatile movements	Masterful instrument use, effective	1	Forward-looking	Strategic and intelligent adjustments	Lack of bimanual dexterity	
PASS NEEDLE THROUGH TWO EDGES OF TISSUE WITH APPROPRIATE BITE PLACEMENT AND TISSUE HANDLING	4	Uncoordinated movements	Stiff and uncontrolled	4	Stagnant	Does not move	Wrong placement to the edge	
	3	Hesitant movement	Controlled, but hesitant and ineffective	3	Lagging	Adjusting with delay or without efficiency	Wrong placement to the next suture	
	2	Skillful movements	Smooth, controlled and meaningful	2	Meaningful	Meaningful adjustment of to improve exposure	Too big or too small bite	
	1	Versatile movements	Masterful instrument use, effective	1	Forward-looking	Strategic and intelligent adjustments	Lack of bimanual dexterity	
CREATE FIRST DOUBLE WIND/THROW OF THE KNOT AND TIGHTEN CORRECTLY	4	Uncoordinated movements	Stiff and uncontrolled	4	Stagnant	Does not move	Failure to create the first double throw	
	3	Hesitant movement	Controlled, but hesitant and ineffective	3	Lagging	Adjusting with delay or without efficiency	Too loose or too tight knot	
	2	Skillful movements	Smooth, controlled and meaningful	2	Meaningful	Meaningful adjustment of to improve exposure	Tightening a twisted knot	
	1	Versatile movements	Masterful instrument use, effective	1	Forward-looking	Strategic and intelligent adjustments	Lack of bimanual dexterity	
KNOT TYING	4	Uncoordinated movements	Stiff and uncontrolled	4	Stagnant	Does not move	Failure to perform a knot	
	3	Hesitant movement	Controlled, but hesitant and ineffective	3	Lagging	Adjusting with delay or without efficiency	Edges are not coapted	
	2	Skillful movements	Smooth, controlled and meaningful	2	Meaningful	Meaningful adjustment of to improve exposure	Knot too tight or too loose	
	1	Versatile movements	Masterful instrument use, effective	1	Forward-looking	Strategic and intelligent adjustments	Lack of bimanual dexterity	
TOTAL SCORE								

Figure 5.2: The CAT form for laparoscopic suturing

	Instruments handling		Tissue handling	
	κ	*p*	κ	*p*
Pickup needle in correct orientation	0.87	< 0.001	0.86	< 0.001
Pass needle through edges of tissue	0.77	< 0.001	0.51	< 0.001
Create first double throw	0.73	< 0.001	0.85	< 0.001
Knot tying	0.75	< 0.001	0.73	< 0.001

Table 5.4: IScore comparisons of the first attempt and the last attempt of the separate LS-CAT items as assessed by the Mann–Whitney U test

Reliability

All participants were able to finish the suturing task. In total, 36 videos of eighteen participants were randomly collected and were scored independently by the two objective observers (observer A and B). Of these participants, seventeen were novices and one was an expert. Mean scores for each separate item are presented in Table 5.1. Cohen's Kappa analysis revealed good to excellent inter-rater agreement scores for almost all tasks of instrument handling and tissue handling (0.87; 0.77; 0.75; 0.86; 0.85, all with $p < 0.001$, Table 5.2). The LS-CAT total scores demonstrated excellent inter-observer reliability for instrument handling ($r = 0.98$, $p < 0.001$), tissue handling ($r = 0.86$, $p < 0.001$), errors ($r = 0.99, p < 0.001$) and the total assessment score ($r = 0.98$, $p < 0.001$). An overview with more detail is presented in Table 5.3.

Performance scores

Within the novice group, subjects performed significantly better at the end of their learning curve compared to their first attempt for all items on the LS-CAT as assessed by both observers. Overall scores are significant for all tasks: instrument handling ($p < 0.001$); tissue handling ($p < 0.001$); pickup needle in correct orientation ($p < 0.001$); pass needle through edges of tissue ($p < 0.001$); create first double throw ($p < 0.001$); knot tying ($p < 0.001$); total amount of errors ($p < 0.001$) and the total assessment score ($p < 0.001$). A full overview of subjects' mean scores and statistics by observer A and B is presented in Table 5.4.

Discussion

Laparoscopic suturing is considered as an essential skill that is required in advanced MIS techniques. Currently, there are no reliable tools that are widely used, to objectively appraise performance in this advanced technique. This is required to influence and promote training and ascertain competency. Mandel et al. already suggested the incorporation of task specific checklist, which has been incorporated in the CAT method with success [11]. The incorporation of this check list was even accurate for self-assessment [14], which is an important finding, because the usability for self-assessment reduces costs and workload for expert instructors [14, 17].

The original concept of CAT has been proven successful to reliably assess technical performance [12]. Based on the method used for the original CAT development, we developed a bespoke laparoscopic suturing competency assessment tool (LS-CAT) that describes and evaluates agreed specific steps in laparoscopic suturing. It evaluates both the process of performance (instrument use, tissue handling and committed errors) and the quality of the end product. Prior to using this new tool in surgical training, multiple criteria must be met, including reliability evidence [4, 18, 19]. This study demonstrated excellent inter-observer reliability for all variables in the adapted CAT form for laparoscopic suturing. Furthermore, a significant difference in performance was found for subject' scores at the beginning and end of their learning curve, indicating the ability of the LS-CAT to discriminate between experience levels within the learning curve.

In the clinical setting, skills are often assessed by experts using the Objective Structured Assessment of Technical Skills (OSATS) form based on the overall performance [14, 20, 21]. However, OSATS do not seem to provide any formative information on the separate skills that still needs to be improved or already is sufficient, which the CAT form does. There is also no clear demonstrated correlation between the OSATS score and outcome of the specific procedure that the resident or surgeon has performed [22], furthermore the trainee does not know which specific skills have to be improved. The scoring of tools like OSATS and its derivatives like the Global Evaluative Assessment of Robotic Skills (GEARS) or generic Global Operative Assessment of Laparoscopic Skills (GOALS) are not specifically designed to provide the information on the separate skills that are being trained.

Other instruments such as a General Rating Scale (GRS) are considered a fair measurement tool, because of the adding of some more specific qualitative assessment parameters (rated on a five-point scale). When using video-recorded performances, this could enhance the objectivity in the ratings of both the OSATS and the GRS; however, these are still not as task specific as the CAT form. Another assessment method often used for surgical skills training (outside the clinical set ting) is motion tracking, which is a highly objective measurement tool used in virtual and augmented reality, and the validity has been proven for numerous systems [19, 23]. However, the quality of the overall task performance might not be assessed sufficiently, because the parameters used are often abstract and not translated to the actual performance of the procedure. Parameters such as 'path length' or 'economy of motion' and 'time' are used, which are not informative of the outcome of the task [24]. These parameters might give an insight in the expertise level of the trainee, but they do not provide information on the accuracy of the task or the final product to indicate competency. Furthermore, a motion tracking system seems to be limited to research centres with available resources, which limits its wider use. The mentioned shortcomings of these assessment methods are not present in the LS-CAT and it requires very little resources and can be generalisable in the assessment and training of laparoscopic suturing skills. Therefore, we think it has the potential as an objective performance assessment for laparoscopic suturing.

Another method for assessment along this model is the Crowd-Sourced Assessment of Technical Skills (C-SATS), which is a type of video assessment performed by large numbers of anonymous online raters [10]. These raters are selfselected from broad sections of the public, thus not every rater may have a medical background. Multiple studies have shown that the inter-observer reliability of a large group of non-expert observers was even better than a smaller group of expert observers for the assessment of surgical performance [25–27] which suggest this method could be used as an assessment tool in surgical technical skills education. The combination of C-SATS with the CAT method could be a powerful mix in terms of time management and cost effectiveness. Both the potential of C-SATS and the usability for self-assessment of the (LS-) CAT form need to be researched in future studies, to fully understand their potential benefits to provide a directive and focused assessment for laparoscopic suturing.

A limitation of this study is that the tool was designed to facilitate categorical qualitative appraisal of skill areas within a series of tasks. Whilst this makes it an effective adjunct to breakdown the task for delivery of constructive feedback on performance, there are certain assumptions that may impact upon its use for summative assessment. There is an assumption that performance in each skill domain and each task is of equal importance (weight) to the overall performance of the procedure. Additionally, the assessment metrics used for the tool were defined by the authors in discussion with experts; however, there may be aspects of performance that were not identified and thus are not evaluated in the current tool. Therefore, other studies are required to validate the tool and clarify its role within the training curriculum for laparoscopic surgery.

CONCLUSION

We developed the LS-CAT, which is a laparoscopic suturing grading matrix to objectively assess the technical performance of laparoscopic suturing, with an excellent inter-rater reliability and the ability to discriminate between experience levels within the learning curve. Although the LS-CAT satisfies many of the requirements of a useful assessment tool with potential application for summative assessment and guide training in this task, further validation studies are required.

REFERENCES

[1] R. Aggarwal, J. Hance, S. Undre, J. Ratnasothy, K. Moorthy, A. Chang, and A. Darzi, *Training junior operative residents in laparoscopic suturing skills is feasible and efficacious,* Surgery **139**, 729 (2006).

[2] J. K. Champion, J. Hunter, T. Trus, and W. Laycock, *Teaching basic video skills as an aid in laparoscopic suturing,* Surgical Endoscopy **10**, 23 (1996).

[3] N. T. Nguyen, K. L. Mayer, R. J. Bold, M. Larson, S. Foster, H. S. Ho, and B. M. Wolfe, *Laparoscopic suturing evaluation among surgical residents,* Journal of Surgical Research **93**, 133 (2000).

[4] K. Ahmed, D. Miskovic, A. Darzi, T. Athanasiou, and G. B. Hanna, *Observational tools for assessment of procedural skills: a systematic review,* The American Journal of Surgery **202**, 469 (2011).

[5] K. Gurusamy, R. Aggarwal, L. Palanivelu, and B. R. Davidson, *Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery,* British Journal of Surgery **95**, 1088 (2008).

[6] K. Moorthy, Y. Munz, A. Dosis, F. Bello, A. Chang, and A. Darzi, *Bimodal assessment of laparoscopic suturing skills: Construct and concurrent validity,* Surgical Endoscopy **18**, 1608 (2004).

[7] S. G. Kroeze, E. K. Mayer, S. Chopra, R. Aggarwal, A. Darzi, and A. Patel, *Assessment of laparoscopic suturing skills of urology residents: A pan-European study,* European Urology **56**, 865 (2009).

[8] S. M. B. I. Botden, I. H. J. T. de Hingh, and J. J. Jakimowicz, *Meaningful assessment method for laparoscopic suturing training in augmented reality,* Surgical Endoscopy **23**, 2221 (2009).

[9] S. Yamaguchi, D. Yoshida, H. Kenmotsu, T. Yasunaga, K. Konishi, S. Ieiri, H. Nakashima, K. Tanoue, and M. Hashizume, *Objective assessment of laparoscopic suturing skills using a motion-tracking system,* Surgical Endoscopy **25**, 771 (2010).

[10] K. Atesok, R. M. Satava, J. L. Marsh, and S. R. Hurwitz, *Measuring surgical skills in simulation-based training,* Journal of the American Academy of Orthopaedic Surgeons **25**, 665 (2017).

[11] L. S. Mandel, B. A. Goff, and G. M. Lentz, *Self-assessment of resident surgical skills: Is it feasible?* American Journal of Obstetrics and Gynecology **193**, 1817 (2005).

[12] D. Miskovic, M. Ni, S. M. Wyles, R. H. Kennedy, N. K. Francis, A. Parvaiz, C. Cunningham, T. A. Rockall, A. M. Gudgeon, M. G. Coleman, and G. B. Hanna, *Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England,* Annals of Surgery **257**, 476 (2013).

[13] N. Britten, *Qualitative research: Qualitative interviews in medical research,* BMJ **311**, 251 (1995).

[14] S. Ganni, M. K. Chmarra, R. H. M. Goossens, and J. J. Jakimowicz, *Self-assessment in laparoscopic surgical skills training: Is it reliable?* Surgical Endoscopy **31**, 2451 (2016).

[15] J. L. Fleiss, *Statistical Methods for Rates and Proportions (Wiley Series in Probability and Statistics)* (Wiley-Interscience, 1981).

[16] C. Cooper, *Individual Differences, 2Ed (Psychology)* (Routledge, 2002).

[17] M. Nagendran, K. S. Gurusamy, R. Aggarwal, M. Loizidou, and B. R. Davidson, *Virtual reality training for surgical trainees in laparoscopic surgery,* Cochrane Database of Systematic Reviews (2013), 10.1002/14651858.cd006575.pub3.

[18] K. Moorthy, *Objective assessment of technical skills in surgery,* BMJ **327**, 1032 (2003).

[19] P. D. van Hove, G. J. M. Tuijthof, E. G. G. Verdaasdonk, L. P. S. Stassen, and J. Dankelman, *Objective assessment of technical surgical skills,* British Journal of Surgery **97**, 972 (2010).

[20] J. R. Eriksen and T. Grantcharov, *Objective assessment of laparoscopic skills using a virtual reality stimulator,* Surgical Endoscopy **19**, 1216 (2005).

[21] R. Aggarwal, K. Moorthy, and A. Darzi, *Laparoscopic skills training and assessment,* British Journal of Surgery **91**, 1549 (2004).

[22] D. D. Anderson, S. Long, G. W. Thomas, M. D. Putnam, J. E. Bechtold, and M. D. Karam, *Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively,* Clinical Orthopaedics and Related Research® **474**, 874 (2015).

[23] J. T. Berlage, S. M. Botden, M. Schijven, and J. J. Jakimowicz, *Face validity study of the ProMIS augmented reality laparoscopic suturing simulator.* Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare **1**, 182 (2006).

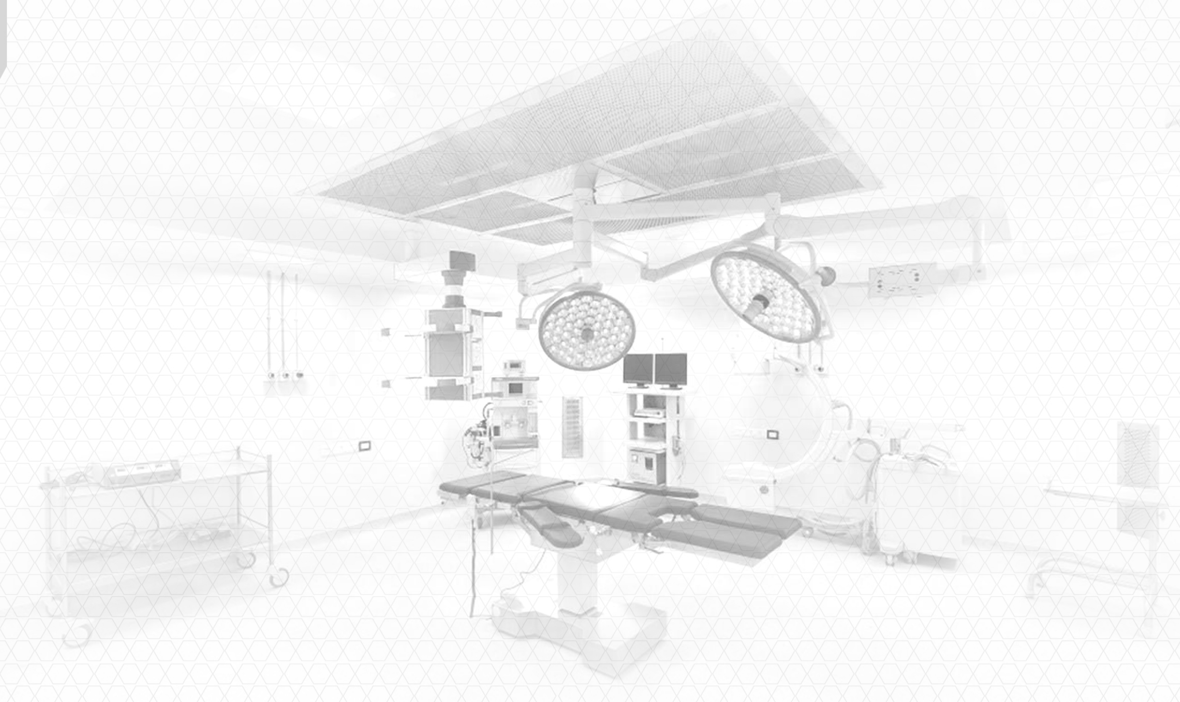
[24] M. C. Porte, G. Xeroulis, R. K. Reznick, and A. Dubrowski, *Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills,* The American Journal of Surgery **193**, 105 (2007).

[25] D. Holst, T. M. Kowalewski, L. W. White, T. C. Brand, J. D. Harper, M. D. Sorenson, S. Kirsch, and T. S. Lendvay, *Crowd-sourced assessment of technical skills: An adjunct to urology resident surgical simulation training,* Journal of Endourology **29**, 604 (2015).

[26] S. B. Deal, T. S. Lendvay, M. I. Haque, T. Brand, B. Comstock, J. Warren, and A. Alseidi, *Crowd-sourced assessment of technical skills: an opportunity for improvement in the assessment of laparoscopic surgical skills,* The American Journal of Surgery **211**, 398 (2016).

[27] D. Holst, T. M. Kowalewski, L. W. White, T. C. Brand, J. D. Harper, M. D. Sorensen, M. Truong, K. Simpson, A. Tanaka, R. Smith, and T. S. Lendvay, *Crowd-sourced assessment of technical skills: Differentiating animate surgical skill through the wisdom of crowds,* Journal of Endourology **29**, 1183 (2015).

6



A SOFTWARE-BASED TOOL FOR VIDEO MOTION TRACKING IN THE SURGICAL SKILLS ASSESSMENT LANDSCAPE

S.Ganni, S.M.B.I.Botden, M.K.Chmarra, R.H.M.Goossens, J.J.Jakimowicz

Published in Surgical Endoscopy, 32, 2994-2999, January 2018

6

Abstract

Background

The use of motion tracking has been proved to provide an objective assessment in surgical skills training. Current systems, however, require the use of additional equipment or specialised laparoscopic instruments and cameras to extract the data. The aim of this study was to determine the possibility of using a software-based solution to extract the data.

Methods

6 expert and 23 novice participants performed a basic laparoscopic cholecystectomy procedure in the operating room. The recorded videos were analysed using Kinovea 0.8.15 and the following parameters calculated the path length, average instrument movement and number of sudden or extreme movements.

Results

The analysed data showed that experts had significantly shorter path length (median $127cm$ vs. $187cm$, $p = 0.01$), smaller average movements (median $0.40cm$ vs. $0.32cm$, $p = 0.002$) and fewer sudden movements (median 14.00 vs. 21.61, $p = 0.001$) than their novice counterparts.

Conclusion

The use of software-based video motion tracking of laparoscopic cholecystectomy is a simple and viable method enabling objective assessment of surgical performance. It provides clear discrimination between expert and novice performance.

Introduction

Current training and evaluation in laparoscopic surgery require a combination of knowledge-based and technical skills assessment [1, 2]. Acquiring the necessary skills takes time, patience and technical aids, such as box trainers, virtual and augmented reality simulators [3–6]. One of the aims of these simulators is the attempt to reduce the reliance upon subjective expert observers when evaluating performance or assessing the acquisition of technical skills [7–10]. This is achieved by motion tracking of the instruments during the performance of, for example, laparoscopic cholecystectomy tasks and procedures in these simulated settings.

Motion tracking is a process where the location, movements, speed and/or acceleration of the instruments used by a surgeon are measured continuously whilst performing a procedure. Current tracking systems use different technologies (e.g. mechanical, optical, acoustic or electromagnetic) to collect the data about the instrument movements and forces applied. The instrument movements and applied forces are the parameters which are used to assess the performance by comparing them against a set of predetermined criteria [11].

The difficulty of extracting a set of criteria suitable for reliable, objective assessment of performance has been a significant challenge for these technical methods [12]. The specific difficulty is how to convert the measures recorded, including instrument position, path length, jerk index, speed, acceleration, etc., into a set of objective criteria which differentiates between competence and weaknesses. It has been shown that motion tracking can in fact be used to generate an objective set of criteria; this, however, necessitates a validation of expert performances to determine the benchmark for optimal performance. [13–16]

Current instrument tracking systems need additional equipment, or the use of special laparoscopic instruments, to facilitate the data acquisition and processing. Including additional recording equipment has not only the disadvantage of cost, but it can be very difficult to use in the clinical setting. Therefore, currently, assessments based on motion tracking are often done outside the clinical setting in a simulator (e.g. augmented or virtual reality simulator). Moreover, these methods only facilitate prospective analysis of surgical procedures and as such cannot be used for retrospective analysis. It would, however, be useful to have an automated tracking system to objectively assess the surgical skills in the clinical setting. Ideally, this tracking system would not require changing the currently used surgical instruments, analyse video recordings of procedures or selected component tasks without requiring pre-preparation, using little to none valuable space in the operating theatre and not hindering the ergonomics and safety of an already difficult form of surgery.

Thus, the aim of this study is to evaluate whether or not it is possible to extract a set of objective criteria from videos of conventional laparoscopic cholecystectomy, by means of motion tracking. For this, a dedicated software has been used to avoid the use of additional equipment in the operating room.

MATERIALS AND METHODS

PARTICIPANTS

All participants recruited for the study were either from Catharina Hospital, Eindhoven, The Netherlands (expert laparoscopic surgeons, who have performed more than 300 laparoscopic procedures) or the participants of the Laparoscopic Surgical Skills Curriculum Grade 1 Level 1 (surgical residents, who have performed fewer than ten laparoscopic procedures). There were 6 expert (over 300 procedures conducted) participants and 23 novice participants. All participants gave their consent for the video recording of them conducting the procedure to be used in this study and hospital ethics approval was obtained.

TASK DESCRIPTION

The participants performed a basic laparoscopic cholecystectomy procedure in the operating room. For both expert and non-expert participants, all the patients operated on were uncomplicated cases without any contraindications. Non-expert participants who needed help from their instructor were excluded from the study. After the procedure, the videos were collected from the operating complex database.

DATA EXTRACTION

Raw video files of the clinical laparoscopic procedures were imported into Kinovea 0.8.15. Initial starting points were identified for three measurements: the first joint of the instrument, and two perpendicular lines, which were used for scale (see Figure 6.1). Coordinate data (x, y) were extracted using the software in semi-automatic mode—that is, where a tracking point (or line) is placed and then the software attempts to

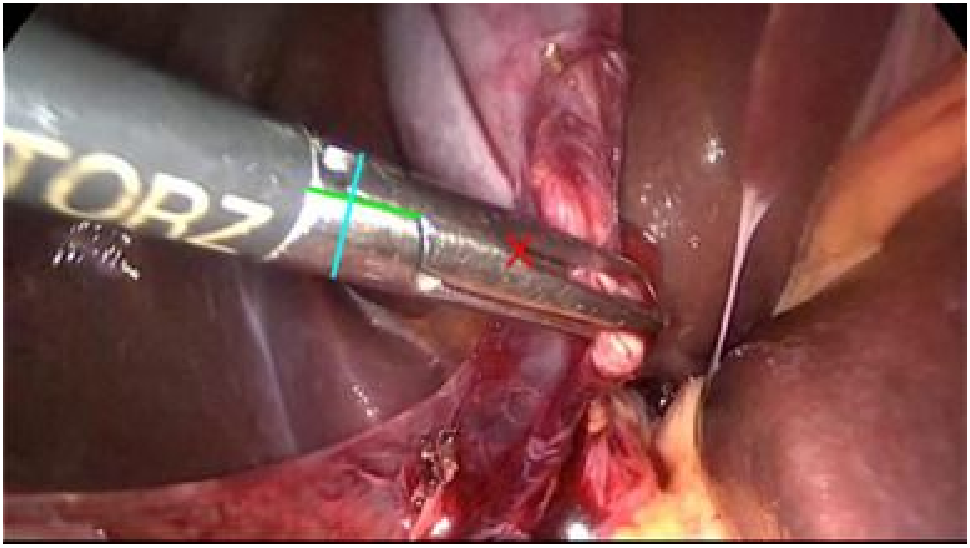


Figure 6.1: Image generated by Kinovea showing the placement of both the starting point (red cross) and the scale-lines (green and blue)

automatically track where the same pixels are in the next frame. This placement was then manually checked and the locations adjusted in the cases where the software had not been able to correctly locate the same point(s) in consecutive frames.

CAMERA DISTANCE

The videos were recorded with a laparoscopic camera during real laparoscopic procedures. During these procedures, an assistant operated the camera. Therefore, it was necessary to scale the (x, y) coordinate data using the (fixed, known) size of the surgical implement in the images. This was achieved by using the two perpendicular lines oriented along the second segment of the implement, from which, being of known length and angle within the image, an $x-$ and $y-$scaling is calculated. This makes videos taken at different distances from the site of the operation easily comparable with each other. The coordinates were then adjusted using this scaling and converted to real-life distances based upon pixel width in the video ready for the calculation of statistics (Table 6.1). The limitation of this scaling is that the angle of the implement in the $z-$direction, may affect the result in a 2D video. This was overcome by using a pair of perpendicular lines for scaling.

Description	Symbol	Formulae	Units
x-coordinate at frame n	x_n		Pixels
y-coordinate at frame n	y_n		Pixels
Number of frames	N		n/a
Number of frames per second	f		n/a
Scaled x-coordinate at frame n	x'_n		cm
Scaled y-coordinate at frame n	y'_n		cm
Distance moved between consecutive frames	d_{n+1}	$\sqrt{(x'_{n+1}-x'_n)^2+(y'_{n+1}-y'_n)^2}$	cm
Total time taken	T	$\frac{N}{f}$	S
Total distance travelled	D	$\sum_{n=1}^{N-1} d_{n+1}$	cm
Average distance travelled per frame	$\overline{D}$	$\frac{D}{N-1}$	cm
Average speed	$\overline{S}$	$\frac{D}{T}$	cm/s
Number of extreme movements	M^e_{n+1}	$\begin{cases} 1 & d_{n+1} \geqslant d_{th} \\ 0 & d_{n+1} < d_{th} \end{cases}$	n/a
Number of extreme movements	E	$\sum_{n=1}^{N-1} M^e_{n+1}$	n/a

Table 6.1: Criteria measured in the procedure and details their calculation

6

Missing data points

Occasionally, the instrument had moved out of the field of view of the camera, or tissue was covering it. The software was able to automatically relocate the point in the majority of cases. In the former case, it was more difficult to relocate the same point for tracking as before (causing the wrong pixel to be tracked going forward). The results in this study were generated by making a decision on a case-by-case basis whether to approximate the location of the point (if enough of the instrument was visible to locate it manually) or to not track that section of the video. If the section was not counted, mean movement was considered to have happened between the last tracked point and the newly found point.

Bias

Additionally, to assess the extent to which bias was a factor in this intervention, the statistics were calculated for a section of the videos of the surgical procedure where the instrument was always visible. Here, the algorithm was allowed to run completely automatically and the results compared with the semi-automatic procedure. These results were then compared with the overall results.

STATISTICS

Three parameters were calculated: the path length, that is the total distance the tip of the instrument has travelled during the procedure; the average distance the instrument tip moved per time frame; and the number of extreme movements (defined as more than 1 cm movement per frame). Four other parameters were calculated from the extracted data using MATLAB (R16b), namely, (1) the Euclidian distance between each consecutive pair of points and (2) the average movement; the number of movements of a distance both (3) under, and (4) above a certain threshold. The statistics were presented using Graphpad Prism and, because the data were non-parametric, the Wilcoxon Signed-Rank test was used to calculate significant differences between the assessment scores. A p value of < 0.05 was considered statistically significant.

RESULTS

In total, the data of 29 participants are included in this study, of which 6 were experts and 23 novices. An example of 18 seconds worth of output from the tracking for one instrument is shown in Figure 6.2 with a 3D representation— $x-$ and $y-$coordinates with time—to make the position of the implement clearer.

PROCEDURAL RESULTS

The summary data for path length, average movement and the number of sudden movements are shown in the box-and-whisker plots in Figure 6.3. Experts had significantly shorter path length (median $127cm$ vs. $187cm$, $p = 0.01$), smaller average movements (median $0.40cm$ vs. $0.32cm$, $p = 0.002$) and fewer sudden movements (median 14.00 vs. 21.61, $p = 0.001$) than their novice counterparts. No statistical difference was seen in path length per minute. (median $41.6cm/min$ vs. $43.6cm/min$).

OVERCOMING BIAS

As can be seen from Figure 6.4, the median response was within 5% for the average distance travelled for each group but the spread of the data was increased by a couple of significant outliers. Manual analysis these outliers revealed that these were all caused by the wrong pixel being identified when the instrument re-entered the frame. It was therefore deemed necessary to manually identify the correct points in these cases.

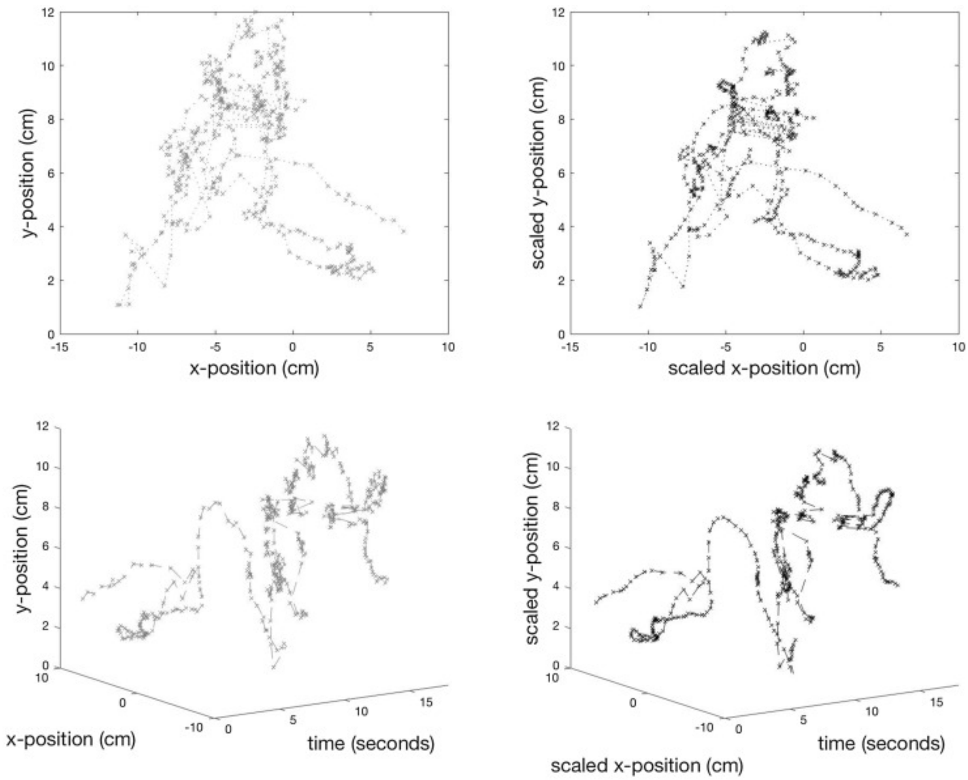


Figure 6.2: A path, unscaled (x, y) for an 18-s period and BOTTOM ROW: the same path scaled for camera movement (x' and y'). LEFT: 2-dimensional visualisation of the measurements; RIGHT: 3-dimensional visualisation (x('), y(') and time) of the measurements

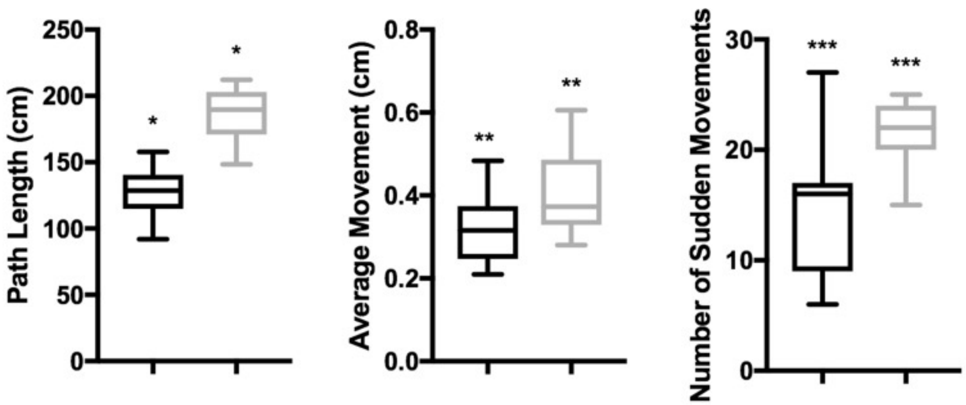


Figure 6.3: Summary data of the total path length, average movement and number of sudden movements (above 3.5 cm). Expert (black) and non-expert (grey) surgeons. *,** and *** indicate pairs of significant difference

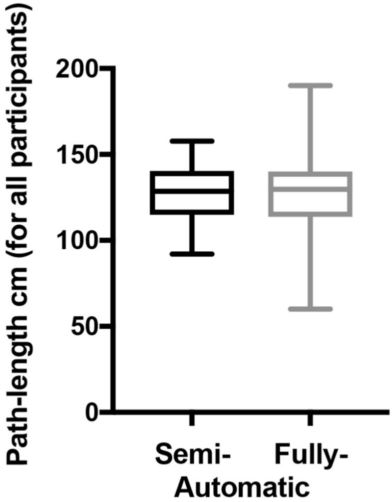


Figure 6.4: Comparison of spread of path length data using semi- (black) and fully automated (grey) procedure

DISCUSSION

The advantages of the ability to evaluate performance during laparoscopic procedures without the need for additional equipment are clear; because, this would allow for an objective clinical assessment. Furthermore, a software-based solution would allow for retrospective evaluation of surgical procedures.

This study aimed to see whether video-based motion tracking system is adequate in differentiating between expert and non-expert outcomes, when their clinical performances are evaluated on video, without any additional equipment installed. The use of this video motion tracking allows for a 2-D, $x - y$ path projection of the 3-D location over time. From this, it is not only possible to extract the time of the procedure, but more importantly the specific movements of the instruments with average speed and number of 'extreme' movements that are made by a surgeon. Our results confirmed our hypothesis that in all three of these criteria, that experts took less time, had more efficient instrument motion and made fewer extreme movements than their non-expert counterparts. No difference was seen in the speed, however, which compliments the results shown in the study conducted by Kowalewski et al. [17]. Taken together, this suggests that it may be possible to discriminate between expert and non- expert participants using this method. It seems that this is the case and thus this system could be used as an alternative to clinically cumbersome and costly methods of motion tracking. This type of retrospective analysis may provide a way for determining the level of performance in laparoscopic surgery in future. However, it is necessary to establish thresholds for safe performance in laparoscopic cholecystectomy. The next step in establishing this would be to determine a set of thresholds/criteria based on the data this study resulted in and validate these in a blinded fashion.

LIMITATIONS

Difficulties and limitations of using video files were overcome in procuring the data that enable objective evaluation of performance. The process is currently only semi-automatic—both in terms of the tracking and, indeed, in deciding the 'window of interest' in terms of the relevant part of the surgical procedure. Furthermore, it was necessary to decide on a case-by-case basis what to do when either the camera's view does not include the instrument's tip (for instance, it is covered by tissue). In particular, it was necessary to consider the effect of the camera's movement in relation to the instrument tip (in all three spatial dimensions) in calculating the average distance moved. Whilst there is no perfect solution to this, our procedure was to use the known size of the instrument's joints in the frame to calculate the relative size of each pixel and then scale by the average of this seems to be a fair compromise. If using this method prospectively, an expert camera driver could be used but, for retrospective use, ideally this process should be automated in the future. In spite of these difficulties, the results clearly discriminate between those procedures performed by expert and novice surgeons.

CONCLUSION

This technical alternative to expert assessment in clinical practise could prove very valuable for the evaluation of surgical skills. Because no extra instruments or additives are needed, this motion tracking system is usable for all surgeons as an objective assessment of skills. The use of video motion tracking of laparoscopic cholecystectomy is a simple and viable method enabling assessment of performance of the procedure. It provides clear discrimination between expert and novice performance.

REFERENCES

[1] R. Aggarwal, K. Moorthy, and A. Darzi, *Laparoscopic skills training and assessment,* British Journal of Surgery **91**, 1549 (2004).

[2] N. Francis, A. Fingerhut, R. Bergamaschi, and R. Motson, eds., *Training in Minimal Access Surgery FIX* (Springer London, 2015).

[3] S. M. B. I. Botden and J. J. Jakimowicz, *What is going on in augmented reality simulation in laparoscopic surgery?* Surgical Endoscopy **23**, 1693 (2008).

[4] S. Bann, A. Darzi, Y. Munz, B. D. Kumar, and K. Moorthy, *Laparoscopic virtual reality and box trainers: is one superior to the other?* Surgical Endoscopy **18**, 485 (2004).

[5] M. P. Schijven, J. J. Jakimowicz, I. A. M. J. Broeders, and L. N. L. Tseng, *The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum,* Surgical Endoscopy **19**, 1220 (2005).

[6] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, and R. M. Satava, *Virtual reality training improves operating room performance,* Annals of Surgery **236**, 458 (2002).

[7] E. Hamilton, D. Scott, J. Fleming, R. Rege, R. Laycock, P. Bergen, S. Tesfay, and D. Jones, *Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills,* Surgical Endoscopy And Other Interventional Techniques **16**, 406 (2001).

[8] K. Gurusamy, R. Aggarwal, L. Palanivelu, and B. R. Davidson, *Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery,* British Journal of Surgery **95**, 1088 (2008).

[9] S. Shanmugan, F. Leblanc, A. J. Senagore, C. N. Ellis, S. L. Stein, S. Khan, C. P. Delaney, and B. J. Champagne, *Virtual reality simulator training for laparoscopic colectomy,* Diseases of the Colon & Rectum **57**, 210 (2014).

[10] I. Oropesa, P. Sánchez-González, P. Lamata, M. K. Chmarra, J. B. Pagador, J. A. Sánchez-Margallo, F. M. Sánchez-Margallo, and E. J. Gómez, *Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery,* Journal of Surgical Research **171**, e81 (2011).

[11] M. K. Chmarra, C. A. Grimbergen, and J. Dankelman, *Systems for tracking minimally invasive surgical instruments,* Minimally Invasive Therapy & Allied Technologies **16**, 328 (2007).

[12] M. K. Chmarra, S. Klein, J. C. F. de Winter, F.-W. Jansen, and J. Dankelman, *Objective classification of residents based on their psychomotor laparoscopic skills,* Surgical Endoscopy **24**, 1031 (2009).

[13] I. Oropesa, M. K. Chmarra, P. Sánchez-González, P. Lamata, S. P. Rodrigues, S. Enciso, F. M. Sánchez-Margallo, F.-W. Jansen, J. Dankelman, and E. J. Gómez, *Relevance of motion-related assessment metrics in laparoscopic surgery,* Surgical Innovation **20**, 299 (2012).

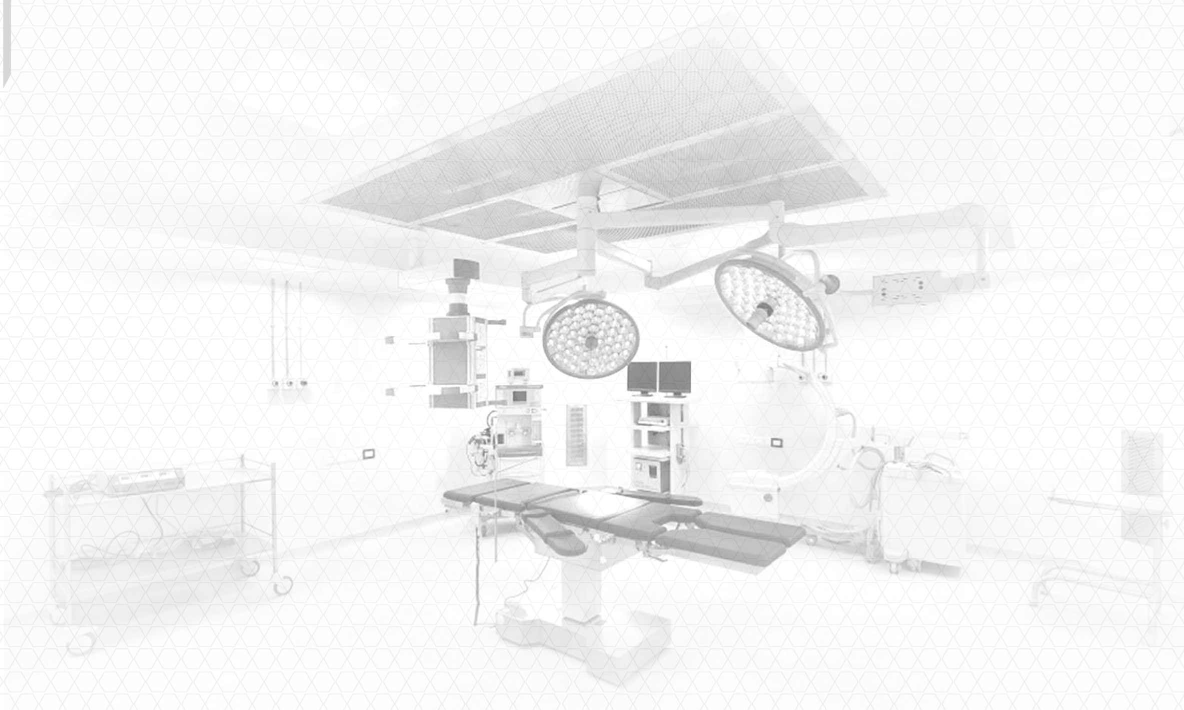
[14] E. F. Hofstad, C. Våpenstad, M. K. Chmarra, T. Langø, E. Kuhry, and R. Mårvik, *A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance,* Surgical Endoscopy **27**, 854 (2012).

[15] A. Ghasemloonia, Y. Maddahi, K. Zareinia, S. Lama, J. C. Dort, and G. R. Sutherland, *Surgical skill assessment using motion quality and smoothness,* Journal of Surgical Education **74**, 295 (2017).

[16] S. Yamaguchi, D. Yoshida, H. Kenmotsu, T. Yasunaga, K. Konishi, S. Ieiri, H. Nakashima, K. Tanoue, and M. Hashizume, *Objective assessment of laparoscopic suturing skills using a motion-tracking system,* Surgical Endoscopy **25**, 771 (2010).

[17] K.-F. Kowalewski, J. D. Hendrie, M. W. Schmidt, C. R. Garrow, T. Bruckner, T. Proctor, S. Paul, D. Adigüzel, S. Bodenstedt, A. Erben, H. Kenngott, Y. Erben, S. Speidel, B. P. Müller-Stich, and F. Nickel, *Development and validation of a sensor- and expert model-based training system for laparoscopic surgery: the iSurgeon,* Surgical Endoscopy **31**, 2155 (2016).

7



VALIDATION OF VIDEO MOTION TRACKING SOFTWARE FOR ASSESSING PERFORMANCE IN LAPAROSCOPIC CHOLECYSTECTOMY

S.Ganni, S.M.B.I.Botden, M.Chmarrayak, M.Li, R.H.M.Goossens, J.J.Jakimowicz

Published in Journal of Medical Systems, 44, 56, January 2020

7

Abstract

Background

Motion tracking software for assessing laparoscopic surgical proficiency has been proven to be effective in differentiating between expert and novice performances. However, with several indices that can be generated from the software, there is no set threshold that can be used to benchmark performances. The aim of this study was to identify the best possible algorithm that can be used to benchmark expert, intermediate and novice performances for objective evaluation of psychomotor skills.

Methods

12 video recordings of various surgeons were collected in a blinded fashion. Data from our previous study of 6 experts and 23 novices was also included in the analysis to determine thresholds for performance. Video recording were analyzed both by the Kinovea 0.8.15 software and a blinded expert observer using the CAT form.

Results

Multiple algorithms were tested to accurately identify expert and novice performances. $\frac{1}{2}L + \frac{1}{3}A + \frac{1}{6}J$ scoring of path length, average movement and jerk index respectively resulted in identifying 23/24 performances. Comparing the algorithm to CAT assessment yielded in a linear regression coefficient R^2 of 0.844.

Conclusion

The value of motion tracking software in providing objective clinical evaluation and retrospective analysis is evident. Given the prospective use of this tool the algorithm developed in this study proves to be effective in bench-marking performances for evaluation.

Introduction

Training and assessment in laparoscopic surgery are increasingly moving towards more objective and criterion-based evaluation tools [1–3]. Box trainers with cameras, virtual and augmented reality simulators have facilitated in achieving objective evaluation of technical skills [4–7]. Recent trends in surgical training, such as self-directed learning and reflective practice, indicate a positive effect of repetitive and independent practice, which have been made possible with objective evaluation tools [8–10]. Several objective criteria such as instrument movement, procedure time, and procedure specific risky manoeuvres can be extracted from these simulators and serve as benchmarks for assessing the performance or self-assessment for progress monitoring [11, 12]. However, the use of these objective criteria in the operating room to assess real surgical procedures is currently limited.

It has been proven by Yamaguchi et al that motion tracking of the surgical instruments can objectively differentiate between expert and novice surgeons in a skills lab setting. This has been achieved using specialized instruments using motion trackers and cameras [13–16]. We have previously used a motion tracking software which is independent of specialized equipment and instruments during the procedure and can be used for retrospective performance analysis using the video recording of the procedure [17]. In this previous study three indices were identified, namely 'path length', 'sudden movements' and 'average movements', which could be extracted from the recorded videos classify expert and novice performances. These indices, however, were procedure specific and as such required a set of benchmarks to assess individual procedures.

Recent advances in image recognition and artificial intelligence (AI) have been proven effective in surgical skills evaluation [18, 19]. These systems are more task and procedure specific, because they evaluate the surgical skills required for laparoscopic knot tying, suturing or pelvic lymph node dissection. But, as with any laparoscopic surgery, skills are broadly categorized into cognitive and psychomotor skills. Cognitive skills as such are procedure specific and psychomotor skills are pan-procedural. Thus, the aim of this study is to develop a new set of benchmarks for psychomotor skills that scale between novice and expert performance and can be used in automated assessment tools.

METHODS

PROTOCOL

To determine a good threshold for the algorithm, the data has to be categorized as shown in Table 7.1. To determine these thresholds, the data from our previous study, [17], was evaluated and recalculated. Three parameters were calculated: 'Path length' (L); 'Average distance' (A), which the instrument tip moved per time frame; and 'Number of extreme movements' (J), defined as more than 1.0 cm movement per frame. If the value of the parameter was above the expert median, a score of 1 was assigned, if it was below the novice median, a score of 0 was assigned. Scores between the two medians were assigned a score between 0 and 1, scaled linearly. Following, these scores were weighted using the following equation, to create a total performance score (p), ranging from 0-1: w_l, w_j and w_a, where $w_l + w_j + w_a = 1$ thus $w_l L + w_j J + w_a A := p$ (Equation 1).

The aim of this study was to calculate the best weightings to determine expertise in the laparoscopic cholecystectomy procedure.

First the original participant data from our previous study was used to determine the expertise thresholds as described above [17]. Following, a blinded evaluation of twelve new videos was performed by both the tracking system and the Competency Assessment Tool (CAT) for laparoscopic cholecystectomy by a blinded assessor to correlate the data. The videos were rated with the new weighting equation and evaluation for a significant correlation. These results were also compared to the previously recorded experience of the surgeon or surgical resident performing the procedure.

7

PARTICIPANTS

These included six experts (>200 laparoscopic procedures performed) and 23 novice participants (<10 laparoscopic procedures performed, but with a surgical background) from the previous study [17]. In addition to this, twelve video recordings of various surgeons and surgical residents, conducting a laparoscopic cholecystectomy at the Catharina Hospital, Eindhoven, The Netherlands, were used to perform a blinded trial to test the accuracy of this thresholding in determining the skill in this procedure. All participants gave their consent for the video recording of the procedures used in this study and hospital ethics committee approval obtained.

Threshold	Category	Procedural requirement
$p >= 2/3$	Expert	201 or more procedures
$1/3 > p < 2/3$	Intermediate	50-200 procedures
$p <= 1/3$	Novice	49 or fewer procedures

Table 7.1: Ideal thresholding output from the algorithm.

Set	Path length (L)	Average Distance (A)	Extreme movements (J)	Correctly Identified
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{20}{24}$
2	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{18}{24}$
3	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{19}{24}$
4	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{15}{24}$
5	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{23}{24}$
6	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{18}{24}$
7	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{21}{24}$

Table 7.2: The values of the weighting parameters for the thresholding and the corresponding number of correctly identified experts and novices.

DATA EXTRACTION AND STATISTICS

7

The tracking data of the instrument movements during the surgical procedure was extracted from the recorded videos using Kinovea 0.8.15 software. Both the thresholding calculations and extracted data were analyzed, including linear regression analysis, using MATLAB (R16b).

RESULTS

THRESHOLD DETERMINED

Data from the tracking software was processed using the thresholding function and Equation described in the methods section, various weightings were evaluated and compared to the correct categorization to identify the best assessment algorithm (Table 7.2).

Set 5 resulted in the most correctly categorized videos, which concluded in the following Algorithm: Assessment score (0-1): Score $= \frac{1}{2}L + \frac{1}{3}A + \frac{1}{6}J$.

VALIDITY OF ASSESSMENT ALGORITHM

Twelve videos were analyzed using the new algorithm with the tracking system and scored using the CAT form by a blinded expert assessor. The thresholding algorithm

Video	Score performance algorithm	Category Identified by thresholds	CAT Score	Actual video category
1	1.00	Expert	21	Surgeon
2	1.00	Expert	22	Surgeon
3	1.00	Expert	20	Surgeon
4	0.86	Expert	19	Surgeon
5	0.67	Expert	20	Surgeon
6	0.63	Intermediate	19	Surgeon
7	0.54	Intermediate	17	Resident
8	0.41	Intermediate	14	Resident
9	0.36	Intermediate	14	Resident
10	0.35	Intermediate	13	Resident
11	0.09	Novice	14	Resident
12	0.00	Novice	13	Resident

Table 7.3: The weighted score is the score calculated using the data extracted for the video and the thresholding equation, performance algorithm. Along with the category that this score yields (from Table 1). The Expert CAT score for that video is also shown and whether the video was, in fact, performed by an experienced surgeon or a student.

categorized the twelve videos as five experts, five intermediates and two novices. The expert-assigned CAT scores support this ordering as shown in Table 7.3. Upon unblinding the data, all the videos identified as expert videos were indeed performed by experienced surgeons and had the top four CAT scores. The other videos evaluated were in fact performances of surgical residents with an intermediate or novice level. Those identified as novices by the algorithm scored the lowest CAT score assigned by the expert assessor. One surgeon was identified as intermediate according to the algorithm, but also scored the lowest CAT score of the surgeons and had a very high jerk index.

7

SIGNIFICANCE LEVEL

The CAT Tool is a comprehensive assessment tool that assesses performance across the three tasks in laparoscopic cholecystectomy in exposure of the cystic duct and artery, cystic pedicle dissection and resection of the gallbladder [20]. These tasks are further evaluated across different indices such as usage of instruments, handling of tissue, errors occurred and the end-product. For this study, we only considered the scoring across the usage of instruments and handling of tissue as they determine the psychomotor skills. Figure 7.1 depicts the linear regression curve plotted using the CAT score and the algorithm yielding a coefficient R^2 of 0.844

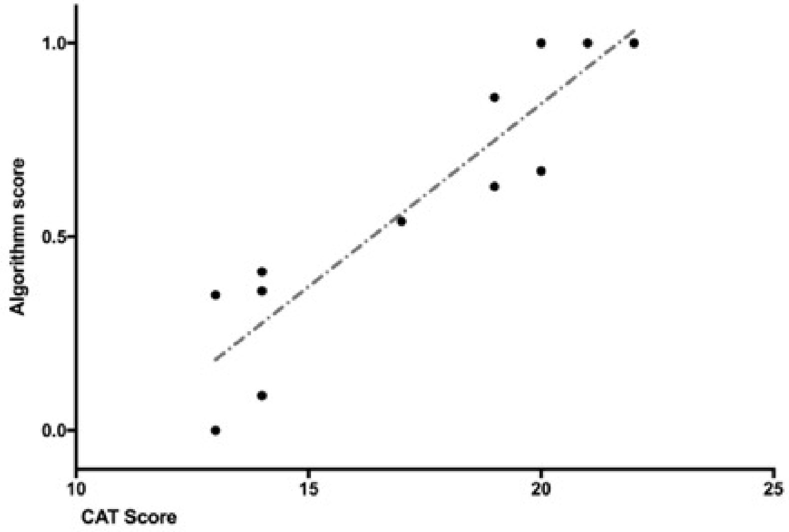


Figure 7.1: Plot of Weighted score of videos, p vs expert-assessed CAT score. The linear trendline has a regression coefficient of determination (R^2) of 0.844

PERFORMANCE SCORING

Scoring systems provide reference for ideal performance and serve as an indicator for measuring learning curve progression and consistency in performance. Upon analysis of the results from the algorithm and correlation with the CAT we propose the following range of scores as derived when using the algorithm for assessing psychomotor skills in laparoscopic cholecystectomy:

- Expert performance: 0.65 and above
- Intermediate performance: 0.35 – 0.65
- Novice performance: 0.35 and below

DISCUSSION

Traditionally assessing surgical skills requires expert assessment through standardized validated tools such as the Competency Assessment Tool (CAT) and Objective Structured Assessment of Technical Skills (OSATS) [20–22]. Objective evaluation of laparoscopic skills using motion analysis has been limited to VR simulators and robotic surgery [23]. The transfer of these evaluation criteria to clinical laparoscopic surgery has been limited by the use of additional equipment and costs [24].

However, this study has shown the potential value of the Kinovea tracking software rapidly evaluate one's performance automatically of a laparoscopic procedure, retrospectively, without the need for additional equipment during the procedure. This new assessment method could be of value in both self-assessments improving the learning curve and as a tool for measuring skills. Because the scoring is by assessing surgical videos retrospectively, there is no bias for the use of other equipment or stress of being watched by an assessor.

Based on the previous study on the feasibility of the Kinovea software [17], the thresholds for the expertise levels were determined using the same data set. This was procedure specific for the laparoscopic cholecystectomy in the clinical setting. The thresholds were set based on a new algorithm, which was validated, by comparing it with both objective expert assessor ($p = 0.01$, $R^2 = 0.844$). Overall, the current threshold algorithm seems to be accurate and a potent objective assessment tool. The algorithm is weighted on the importance of each of the indices identified and the rate in which these make up the expertise of the performance.

Computer vision techniques and AI have shown promising results in identifying procedure specific evaluations [18, 19]. Their strengths lie in detecting cognitive and clinical skills in addition to error recognition. AI can also effectively segment procedural steps for easy access and indexing for future reference [25]. However, these systems do not identify psychomotor skills that can be applied pan procedurally and can serve as an important indicator for learning curve monitoring in the clinical context.

The current calculations used in this study are limited in their application to assessing the skills required for laparoscopic cholecystectomy. Furthermore, these indices cannot provide indication of errors or potential errors. However, with the new insights of this study in the categorization of the importance of performance indices, it could be transferred to other laparoscopic procedures. In combination with computer vision techniques and AI this could serve in providing comprehensive evaluation of laparoscopic skills similar to that of VR simulators in a clinical setting.

CONCLUSION

The value of motion tracking software in providing objective clinical evaluation and retrospective analysis is evident. Given the prospective use of this tool the algorithm developed in this study proves to be effective in benchmarking performances for psychomotor evaluation of laparoscopic skills.

REFERENCES

[1] K. Moorthy, *Objective assessment of technical skills in surgery,* BMJ **327**, 1032 (2003).

[2] P. D. van Hove, G. J. M. Tuijthof, E. G. G. Verdaasdonk, L. P. S. Stassen, and J. Dankelman, *Objective assessment of technical surgical skills,* British Journal of Surgery **97**, 972 (2010).

[3] I. Oropesa, P. Sánchez-González, P. Lamata, M. K. Chmarra, J. B. Pagador, J. A. Sánchez-Margallo, F. M. Sánchez-Margallo, and E. J. Gómez, *Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery,* Journal of Surgical Research **171**, e81 (2011).

[4] S. M. B. I. Botden and J. J. Jakimowicz, *What is going on in augmented reality simulation in laparoscopic surgery?* Surgical Endoscopy **23**, 1693 (2008).

[5] S. Bann, A. Darzi, Y. Munz, B. D. Kumar, and K. Moorthy, *Laparoscopic virtual reality and box trainers: is one superior to the other?* Surgical Endoscopy **18**, 485 (2004).

[6] M. P. Schijven, J. J. Jakimowicz, I. A. M. J. Broeders, and L. N. L. Tseng, *The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum,* Surgical Endoscopy **19**, 1220 (2005).

[7] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, and R. M. Satava, *Virtual reality training improves operating room performance,* Annals of Surgery **236**, 458 (2002).

[8] S. Ganni, M. K. Chmarra, R. H. M. Goossens, and J. J. Jakimowicz, *Self-assessment in laparoscopic surgical skills training: Is it reliable?* Surgical Endoscopy **31**, 2451 (2016).

[9] A. K. Gardner and K. Abdelfattah, *Getting better all the time? Facilitating accurate team self-assessments through simulation,* BMJ Simulation and Technology Enhanced Learning , bmjstel (2019).

[10] S. Ganni, S. M. Botden, D. P. Schaap, B. H. Verhoeven, R. H. Goossens, and J. J. Jakimowicz, *"reflection-,* **75**, 527 (2018).

[11] T. P. Grantcharov, J. Rosenberg, E. Pahle, and P. Fench-Jensen, *Virtual reality computer simulation,* Surgical Endoscopy **15**, 242 (2001).

[12] P. Lamata, E. J. Gomez, F. Bello, R. L. Kneebone, R. Aggarwal, and F. Lamata, *Conceptual framework for laparoscopic VR simulators,* IEEE Computer Graphics and Applications **26**, 69 (2006).

[13] S. Yamaguchi, D. Yoshida, H. Kenmotsu, T. Yasunaga, K. Konishi, S. Ieiri, H. Nakashima, K. Tanoue, and M. Hashizume, *Objective assessment of laparoscopic suturing skills using a motion-tracking system,* Surgical Endoscopy **25**, 771 (2010).

[14] I. Oropesa, M. K. Chmarra, P. Sánchez-González, P. Lamata, S. P. Rodrigues, S. Enciso, F. M. Sánchez-Margallo, F.-W. Jansen, J. Dankelman, and E. J. Gómez, *Relevance of motion-related assessment metrics in laparoscopic surgery,* Surgical Innovation **20**, 299 (2012).

[15] E. F. Hofstad, C. Våpenstad, M. K. Chmarra, T. Langø, E. Kuhry, and R. Mårvik, *A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance,* Surgical Endoscopy **27**, 854 (2012).

[16] A. Ghasemloonia, Y. Maddahi, K. Zareinia, S. Lama, J. C. Dort, and G. R. Sutherland, *Surgical skill assessment using motion quality and smoothness,* Journal of Surgical Education **74**, 295 (2017).

[17] S. Ganni, S. M. B. I. Botden, M. Chmarra, R. H. M. Goossens, and J. J. Jakimowicz, *A software-based tool for video motion tracking in the surgical skills assessment landscape,* Surgical Endoscopy **32**, 2994 (2018).

[18] M. C. Vassiliou, L. S. Feldman, C. G. Andrew, S. Bergman, K. Leffondré, D. Stanbridge, and G. M. Fried, *A global assessment tool for evaluation of intraoperative laparoscopic skills,* The American Journal of Surgery **190**, 107 (2005).

[19] J. A. Martin, G. Regehr, R. Reznick, H. Macrae, J. Murnaghan, C. Hutchison, and M. Brown, *Objective structured assessment of technical skill (OSATS) for surgical residents,* British Journal of Surgery **84**, 273 (1997).

[20] K.-F. Kowalewski, C. R. Garrow, M. W. Schmidt, L. Benner, B. P. Müller-Stich, and F. Nickel, *Sensor-based machine learning for workflow detection and as key to detect expert level in laparoscopic suturing and knot-tying,* Surgical Endoscopy **33**, 3732 (2019).

[21] A. Baghdadi, A. A. Hussein, Y. Ahmed, L. A. Cavuoto, and K. A. Guru, *A computer vision technique for automated assessment of surgical performance using surgeons'*

console-feed videos, International Journal of Computer Assisted Radiology and Surgery **14**, 697 (2018).

[22] D. Miskovic, M. Ni, S. M. Wyles, R. H. Kennedy, N. K. Francis, A. Parvaiz, C. Cunningham, T. A. Rockall, A. M. Gudgeon, M. G. Coleman, and G. B. Hanna, *Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England,* Annals of Surgery **257**, 476 (2013).

[23] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, *Review of methods for objective surgical skill evaluation,* Surgical Endoscopy **25**, 356 (2010).

[24] M. K. Chmarra, C. A. Grimbergen, and J. Dankelman, *Systems for tracking minimally invasive surgical instruments,* Minimally Invasive Therapy & Allied Technologies **16**, 328 (2007).

[25] D. A. Hashimoto, G. Rosman, M. Volkov, D. L. Rus, and O. R. Meireles, *Artificial intelligence for intraoperative video analysis: Machine learning's role in surgical education,* Journal of the American College of Surgeons **225**, S171 (2017).

8

A COMPREHENSIVE THREE-YEAR VALIDATION AND IMPACT STUDY OF A MULTI-MODAL TRAINING CURRICULUM ON LAPAROSCOPIC SURGICAL SKILLS TRAINING

S.Ganni, S.M.B.I.Botden, S.Nayak, B.R.Ganni, R.H.M.Goossens, J.J.Jakimowicz

Submitted to Journal of Surgical Education

8

Introduction

Minimal Access Surgery (MAS) has seen a paradigm shift ever since its introduction in the early 1980s compared to other disciplines of medical training. [1–3] This can be attributed to the unique set of tools and techniques required to perform MAS. Tools that consist of laparoscopic instruments, cameras, energized instruments which are technology driven and dependent and subject to rapid advances and changes at frequent intervals. [4, 5] The skills required to perform are equally challenging in that the surgeon has to cope with the usage of the above-mentioned equipment, hand-eye coordination, 2D representation of the anatomy, fulcrum effect, restricted movement, difficult ergonomics and limited visual field. [6–9]

Contrary to the advances, training in MAS has seen a variegated and often limited to traditional training approach in several parts of the world. [10] Traditional training or the apprenticeship model has been a mainstay for several modalities of training often clouted with the cult of the individual rather than evidence-based learning. [11] Such training models often fall under Halsted's 'see one, do one, teach one' approach. [12] The slow progression of a trainee begins with assisting the expert, performing a procedure with expert assistance and performing a procedure without assistance. [13] MAS training as such with its multi-disciplinary skill set does not bode well with apprenticeship training where systematic and sequential learning is hindered. [14]

Few notable educational theories such as Kolb's learning cycle and Gardner's multiple intelligence theory have been adapted to the MAS training curricula. Kolb's educational theory focuses on the process of learning wherein it is divided into experiencing, reflection, conceptualization and planning. Experiencing the process of doing, Reflection on what has been done, Conceptualizing the process and deciphering the event and Planning the subsequent outcomes. [15] Whereas Gardner's multiple intelligence theory is realized in the learning style; more specifically bodily kinesthetics and spatial intelligence that translates to achieving efficient psychomotor skills crucial for surgical trainees. [16]

Additionally, several training tools are available for surgeons to cope with the challenges of MAS that include box trainers, augmented reality and virtual reality trainers. [17–20] These tools aid in providing a repetitive and objective evaluation of performance over time. The metrics used within range from basic psychomotor skills to advanced comprehensive procedural evaluations. [21–23] The culmination of evidence-based educational theories and modern training tools is the foundation of current curricula in MAS. [24–27] However, implementation of these curricula requires significant infrastructure pertaining to skills training labs, expert faculty and expensive equipment. [10, 28] Several other factors impede the implementation of training in MAS, including but not limited to educational and government regulations, industry support, financial limitations, and increased resident working hour restrictions contrary to the increase in equipment, techniques and procedures they have to cope with. [29–31] Furthermore, socio-cultural barriers, ancillary crew training and readiness to adapt have a significant impact on adapting the latest training methodologies. [10]

In this study, we observe the impact of The Laparoscopic Surgical Skills (LSS) Curriculum over a three-year period on a group of surgical trainees in comparison to the apprenticeship training program in India. Because simulation-based training is still in its early phases of implementation across the country, it has enabled us to study participants in two completely different forms of training independent of each other. Additionally, this study aims to evaluate the direct and indirect benefits of both the training modalities in respect to patient, preventive and hospital outcomes in the hope that it provides a detailed insight and overview for all stakeholders interested in adopting such multi-modal curricula.

METHODS

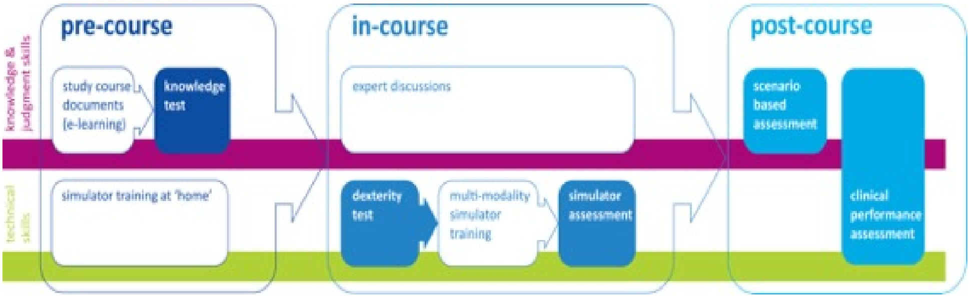


Figure 8.1: The Laparoscopic Surgical Skills Curriculum

PARTICIPANTS

Forty-eight surgical residents six months into their PGY-1 were asked to participate in the study. The test group consisted of twenty-four surgical residents who were from GSL Medical College, India with access to a well-equipped surgical skills lab and where the LSS curriculum was implemented along with their regular residency program. The control group consisted of twenty-four surgical residents from Rangaraya Medical College, Kakinada, India who were following the apprenticeship training program.

INCLUSION CRITERIA

All surgical residents from the test and control group had to pass a knowledge-based questionnaire that is a standard admission test for the LSS Grade 1 Level 1 curriculum. Consequently, they were given access to the reading material that elaborated on the basics of laparoscopy that included ergonomics, positioning, instrumentation and laparoscopic cholecystectomy. This was to ensure all the participants had adequate cognitive knowledge prior to their respective training programs and were in similar academic competence.

8

THE LSS CURRICULUM

The LSS curriculum is a comprehensive multi-level training program ranging from basic to advanced specialty procedures. The Grade 1 Level 1 curriculum is the preliminary program that is aimed at surgical residents and those surgeons new to laparoscopy. [26] The curriculum is divided into three components: Pre-course knowledge test, Two-day course and assessment and Post-course assessment. The outline of the course is depicted in Figure 8.1.

Figure 8.2: From left to right: Pattern cutting, Ball transfer and Block building

Pre-course knowledge test When a participant enrols into the course he/she is assigned a unique identifier that is used throughout the program until certification. This ID allows the participant online access to a recommended list of readings that are evidence-based and cover the basics of laparoscopy and laparoscopic cholecystectomy and appendectomy. The participant has a three-week period to prepare and undertake the online knowledge test before attending the two-day LSS course.

Two-day course After passing the pre-course test the participants are admitted into the two-day course set in a clinical skills lab equipped with box trainers, virtual reality simulators and wet lab facility. The course is scheduled with interactive closed group discussions on the basics of laparoscopy and case studies on laparoscopic cholecystectomy and appendectomy. The participants are rotated to practice on box trainers equipped with an inbuilt camera and light for psycho-motor skills development. The residents were asked to perform ball transfer into an endo-bag, cutting spread gloves in a circular manner, peeling grapes and building blocks as shown in Figure 8.2.

The wet lab training included crucial task training on trocar and port insertion on a section of pig abdominal wall, suturing exercises on intestinal walls and vascular suturing exercises on aortas. Expert mentors were made available to demonstrate these tasks prior along with video presentations of tasks during practice. Mentors were also present to monitor progression and provide feedback when needed. The virtual reality (VR) simulators used in the study were LapMentor™ from Simbionix, USA. The simulator is height-adjustable with a monitor, keyboard and trackball. The procedures are performed using the two instrument handles that replicate real laparoscopic handles along with a camera handle in a trocar and foot pedals as shown in Figure 8.3. The instruments can be interchanged across the range of laparoscopic instruments depending on the task. These are equipped with realistic haptic feedback that adapts to

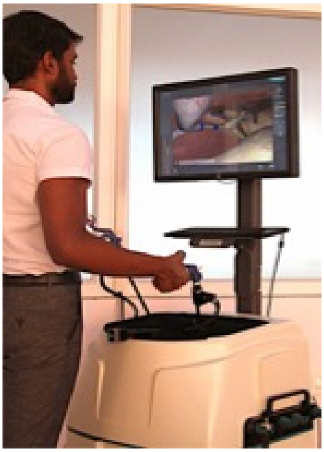


Figure 8.3: Setup of The Simbionix LapMentor

the tissue being handled. The unique ID given to the participants can be used to login to the simulator where they can monitor their progression and task list. Participants were briefed on the usage of simulator until they were familiar. An initial dexterity test with threshold scores was needed for the participants to continue in the course. The dexterity test consisted of camera manipulation, peg transfer and pattern cutting as shown in Figure 8.4.

After the dexterity test, they progressed to component tasks of laparoscopic cholecystectomy and laparoscopic appendectomy. These included identifying cystic duct and artery and demonstrating a critical view of safety, clipping and cutting the cystic duct and artery, removal of the gallbladder from the liver bed. Each component task had a threshold time and performance index the participant aimed to achieve prior to evaluation.

In-course assessment Participants were subject to three assessments during the two-day period. A complete laparoscopic cholecystectomy procedure on the VR simulator and on a porcine liver model and scenario-based assessment at the end of the course on a computer using the Competency Assessment Tool (CAT) designed for laparoscopic cholecystectomy. [32] The CAT is an operation-specific assessment with evaluation criteria spread across three procedural tasks: exposure of cystic artery and cystic duct, cystic pedicle dissection and resection of gallbladder from the liver. Within these tasks, the performance was rated on a five-point task-specific scale based on the usage of instruments, handling of tissue with the non-dominant hand (NDH), errors within each task and the end product of each task. Before the assessments, the

Figure 8.4: Camera manipulation and Peg transfer tasks on the Simbionix LapMentor

participants were informed of the benefits of self-assessment and the usage of the assessment tool with examples of good and poor performances where applicable via pictures and video sources. In addition, the participants who were assisting the assessee were also asked to peer-assess their performance to inculcate an objective perspective. After the procedure the expert assessor and the assessee have a debriefing session that includes comparison of self and expert assessed scores and formative feedback where appropriate.

At the end of the course the participants undertook the scenario-based assessment. This assessment is a computer-based test that test knowledge and judgement skills on index procedures. Participants are presented with different scenarios that test their situation-based judgement on instrumentation and patient condition. Passing the three in-course assessments is crucial for clinical assessment.

Post-course clinical assessment After the two-day course and assessments the participants are allowed to perform an uncomplicated laparoscopic cholecystectomy procedure under the guidance of an expert surgeon. The participant was assessed using Global Assessment Score (GAS) for laparoscopic cholecystectomy. [33] This assesses the participants ability to setup the operating room (OR) for the procedure and in each component task of laparoscopic cholecystectomy. In case of any complication or inability to progress further, the expert surgeon overtook the procedure and was recorded as incomplete. For the purpose of this study, procedures 1 and 4 were noted to observe skills progression and acquisition. After successful completion of four procedures the participants were awarded the LSS Grade 1 Level 1 certificate.

The Apprenticeship Training Program

In this program, the residents were trained as part of the prescribed surgical residency curriculum. In the OR, the resident is responsible for coordinating the patient preparation, instrumentation and paperwork. During the procedure the resident assists the operating surgeon by either holding the camera, handling the instrument transfer, paperwork and documentation. Furthermore, the resident is responsible for the post-operative closure of incision sites, follow-up, instructions for care and medication or dressing where applicable. After a few months into the program, the resident is offered to perform a procedure with the expert surgeon assisting. The expert surgeon guides and when necessary takes over the procedure. Similar to that of the LSS group the expert surgeons were asked to assess the participants to evaluate the residents using the GAS tool. The video recordings of the performances were collected for further evaluation.

Evaluation of Performance

In addition to the GAS scores obtained from both the LSS and Apprenticeship group, several other tools were used to assess the overall performance. Firstly, the first and fourth video recordings were assessed by two expert assessors using the CAT tool. The inter-rater reliability was consistent and they were blinded to the groups training status. After, the videos were analysed using a motion tracking software designed to identify psychomotor skills. [34] The videos were then further analysed using an algorithm that could indicate the proficiency of performance with clear indication of expert, novice and intermediate scores.

Other criteria that were observed apart from surgical performances were, total operating time, surgical disposables used, post-operative complications if any, duration of patient admission and discharge, average cost per patient, treatment turn-over rate and morbidity related indirect expenses.

Results

Participants

Initially, each group had a total of 24 participants. However, 2 and 6 were excluded from the study and control groups, respectively due to not completing the required four procedures. This left a study group of 22 and a control group of 18.

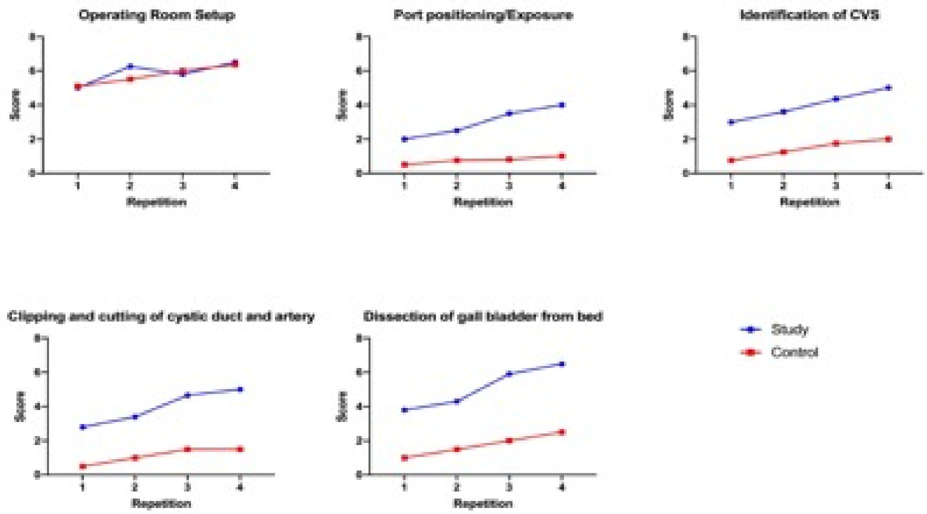


Figure 8.5: Mean GAS results with SEM for each of the four procedures n=1-4 with the control group shown in red and study group in blue. All points are statistically different from the above.

GAS RESULTS

Results from the GAS scores seen in Figure 8.5, indicate that the study group starts with a higher GAS score across all aspects and this gap continues to be present as experience is gained between procedure 1 and 4. Further, the control group scores appear to show more linear progress between procedures whereas the study group scores appear to have more sigmoidal progress. Figure 8.6 shows that the difference between each group for each category at procedure 4 - we see that each study group result is significantly different from the control group.

PROPORTION OF COMPLETIONS, TIME TO COMPLETE AND COMPARATIVE COST

After 4 procedures all of the LSS groups successfully completed the procedure without the need for an experienced surgeon to intervene. After 4 procedures only 8 out of 18 (44.4%) participants managed to complete the operation (Figure 8.7). The total time taken to complete the procedure was a factor considered.

If a participant was unable to complete an operation this could be for a variety of reasons, including a complication, wound infection or bile leakage or surgeon-error which can result in an increased hospital stay, and thus increased cost. Figure 8.8a shows the cost of the procedure (without hospital stay costs). Figure 8.8b, showing the

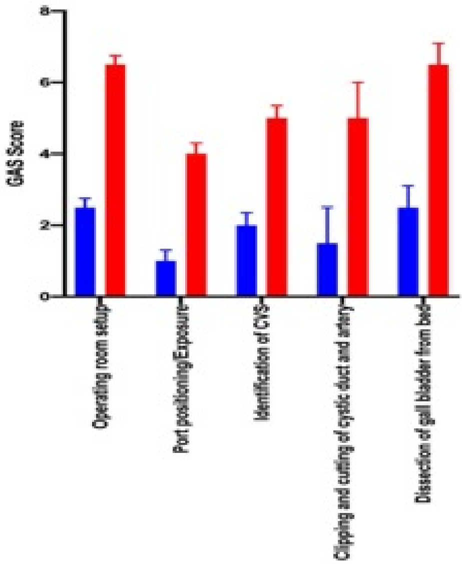


Figure 8.6: GAS score for each category at procedure 4 with the control group shown in red and study group in blue.

Description	**Quantity**	**Cost**
Surgeon	Per procedure	Rs.10,000
Anesthetist	Per 1 hour	Rs.2,500
	Each extra 30 minutes	Rs. 1,000
Operating room	Per 1 hour	Rs.6,000
(Including oxygen, instruments and facility)	Each extra 30 minutes	Rs.3,000
Hospital stay	Per day	Rs.2,000

Table 8.1: The cost of hospital resources.

total cost of the procedure including any hospital stay needed afterwards was constructed using the data in Table 8.1.

MOTION TRACKING ASSESSMENT

Performing motion tracking automatic assessment for all procedures which were completed by the participant, resulting in the scores shown in Table 8.2. All participants were identified as novices in procedure 1. By contrast, two members of the study group were identified as Intermediate in procedure 4. There is a statistically significant improvement in the paired scores between procedure 1 and 4 in the study group ($p < 0.0001$, $n = 18$) but not in the control group ($p = 0.0625$, $n = 6$) (Figure 8.9).

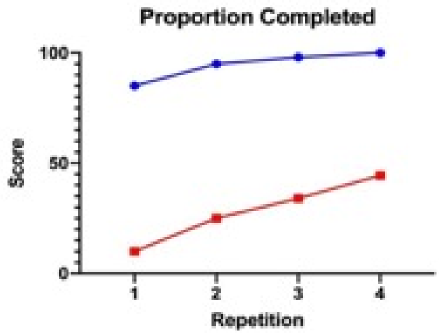


Figure 8.7: Proportion of people who completed the procedure without the intervention of the resident. Study group (blue) and control (red).

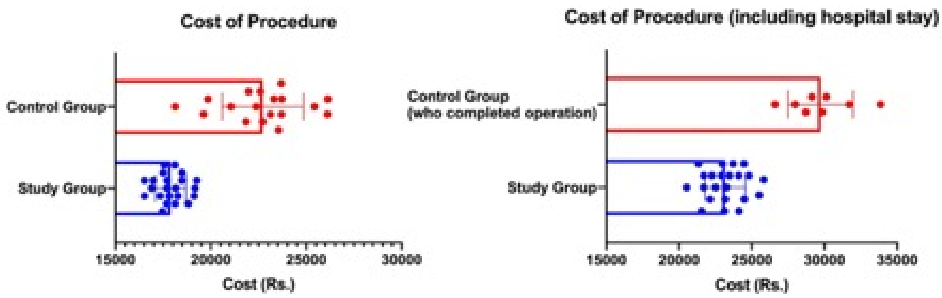


Figure 8.8: Cost of each procedure in terms of a) the cost of surgeon, operating theatre and anesthetist's time b) also including any hospital stay costs.

8

Agreement is seen in the correlation between the CAT scores and the category identified by thresholding ($r^2 = 0.86$).

Discussion

The benefits of novel training curricula and simulators have been documented with great success in the past few decades. [35, 36] They show successful transfer of skills from training environments to operating rooms. [37, 38] However, the impact of such courses on participants over time is limited. This study addresses the impact of a multi-model training curriculum over the course of a surgical trainees training period in contrast to a control group of surgical trainees who had no exposure to a formal skills lab training and were trained in a traditional apprenticeship setting.

Traditional and new tools for evaluation of surgical expertise were used such as the GAS in the OR, CAT and motion tracking assessment after the procedure. These were

Procedure 1	Procedure 4					
Video	Score performance algorithm	Category Identified by thresholds	CAT	Score performance algorithm	Category Identified by thresholds	CAT
1	0	Novice	18	0.31	Novice	22
2	0	Novice	19	0.4	Novice	24
3	NC			NC		
4	NC			NC		
5	NC			0	Novice	20
6	NC			NC		
7	0.06	Novice	20	0	Novice	25
8	0	Novice	21	NC		
9	NC			NC		
10	NC			0.23	Novice	22
11	NC			NC		
12	0.06	Novice	21	0.13	Novice	21
13	0.21	Novice	24	0.42	Novice	28
14	NC			NC		
15	NC			0.22	Novice	24
16	NC			NC		
17	0.13	Novice	24	0.21	Novice	27
18	NC			NC		3
1	0.21	Novice	26	0.61	Novice	33
2	NC			0.18	Novice	23
3	0	Novice	21	0	Novice	23
4	NC			0.23	Novice	30
5	NC			0.5	Novice	32
6	0.04	Novice	20	0.37	Novice	33
7	0	Novice	24	0.3	Novice	30
8	0	Novice	25	0.23	Novice	31
9	0	Novice	25	0.03	Novice	22
10	0.33	Novice	33	0.73	Intermediate	36
11	0	Novice	24	0.25	Novice	31
12	0	Novice	24	0	Novice	20
13	0.41	Novice	28	0.68	Intermediate	35
14	0	Novice	25	0.23	Novice	29
15	0.36	Novice	27	0.59	Novice	32
16	0	Novice	23	0.05	Novice	27
17	NC			0.05	Novice	24
18	0.55	Novice	31	0.63	Novice	33
19	0.22	Novice	24	0.62	Novice	31
20	0.44	Novice	27	0.77	Intermediate	32
21	0.36	Novice	28	0.79	Intermediate	31
22	0	Novice	19	0.18	Novice	28

Table 8.2: Table showing the output of the motion-tracking algorithm for Procedure 1 and 4 for each member of the control group and study group. Those procedures which were not completed are labelled NC.

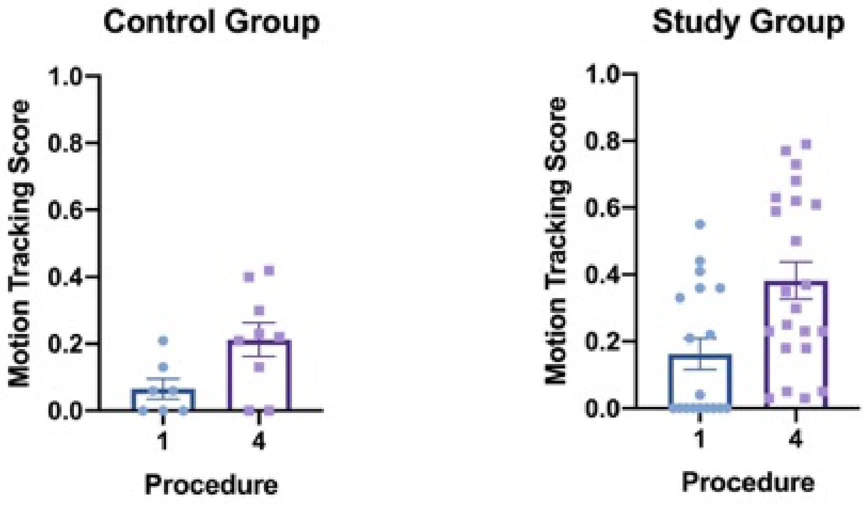


Figure 8.9: Motion tracking scores for the control group and study group in procedures 1 and 4.

documented for the first and fourth procedure to observe progression of skills acquisition and proficiency. The acquisition of skills for the study group was significantly better than the control group in almost all criteria except for non-technical skills such as operation room set up. This could be attributed to the prolonged period of time the control group spent in the OR assisting the surgeon prior to the procedure. The acquisition of skills in both the groups progressed from the first to fourth procedure, however, it was linear in the study group and sigmoidal in the control group. This could be due to lack of proper hands on training for the control group who then had to perform the procedure. The rate of acquisition peaked from the first to second procedure in the control group but the progression from then on was less than that of the study group. In contrast, the study group's rate of acquisition started off with an initial high score and progressed in a linear mode indicating confident learning and progressing to intermediate and expert skill levels. In addition, the rate of non-completion of procedure was significantly higher in the control group (80%) through the fourth procedure in contrast to study group (20%).

The motion tracking software used in the study has been proven to differentiate between novice, intermediate and expert performances by using specific thresholds. [34] Using the software to analyze the performance, we could see that though trainees of both groups were identified as a novice in the first performance; study group trainees progressed to intermediate performance by the fourth procedure and none in the control group.

Considering the cost of procedure between the two groups, significant differences were observed due to increased duration of procedure, use of additional disposable equipment and further extended stay in the hospital. When investigating the reason for extended stay per procedure, the control group trainees attributed it to lack of confidence in the procedure performed and thus extended monitoring. In addition, individual case complications like gall leakage or injury to the gallbladder, cystic duct injury, wound site infection were among the factors contributing to extended hospital stay.

Multi-modal curricula such as the LSS curriculum used in this study can not only successfully aid in transferring skills acquired in training to the OR, but also help in progression of skills acquired therein. Though surgical trainees trained under traditional apprenticeship model perform satisfactorily considering the lack of such structured curricula, the rate of skills acquisition and progression is significantly slower and often gained through learning from errors and patient morbidity.

Current technology and equipment have enabled the widespread use of training curricula that offer repetitive practice and objective evaluation. Despite the initial capital cost of these skills labs and tools the surgical trainees, the institution they practice in and the patients benefit greatly in the long run.

REFERENCES

[1] A. Darzi and S. Mackay, *Recent advances in minimal access surgery,* BMJ **324**, 31 (2002).

[2] J. C. Rosser, M. Murayama, and N. H. Gabriel, *Minimally invasive surgical training solutions for the twenty-first century,* Surgical Clinics of North America **80**, 1607 (2000).

[3] E. B. Bass, H. A. Pitt, and K. D. Lillemoe, *Cost-effectiveness of laparoscopic cholecystectomy versus open cholecystectomy,* The American Journal of Surgery **165**, 466 (1993).

[4] A. Cuschieri, *Technology for minimal access surgery,* BMJ **319**, 1304 (1999).

[5] G. Berci, *Endoscopy* (Appleton-Century-Crofts, 1976).

[6] T. G. Frank, G. B. Hanna, and A. Cuschieri, *Technological aspects of minimal access surgery,* Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine **211**, 129 (1997).

[7] M. A. van Veelen, E. A. L. Nederlof, R. H. M. Goossens, C. J. Schot, and J. J. Jakimowicz, *Ergonomic problems encountered by the medical team related to products used for minimally invasive surgery,* Surgical Endoscopy **17**, 1077 (2003).

[8] G. Lee, T. Lee, D. Dexter, C. Godinez, N. Meenaghan, R. Catania, and A. Park, *Ergonomic risk associated with assisting in minimally invasive surgery,* Surgical Endoscopy **23**, 182 (2008).

[9] R. Berguer, D. L. Forkey, and W. D. Smith, *Ergonomic problems associated with laparoscopic surgery,* Surgical Endoscopy **13**, 466 (1999).

[10] S. Ganni, S. M. B. I. Botden, B. F. Hamilton, A. S. Bedi, D. Lomanto, B. R. G, and J. J. Jakimowicz, *Current state of training and evaluation of laparoscopic surgical skills,* Journal of Surgery: Open Access **3** (2017), 10.16966/2470-0991.148.

[11] J. M. Hamdorf and J. C. Hall, *Acquiring surgical skills,* British Journal of Surgery **87**, 28 (2000).

[12] S. V. Kotsis and K. C. Chung, *Application of the "see one, do one, teach one" concept in surgical training,* Plastic and Reconstructive Surgery **131**, 1194 (2013).

[13] J. M. Hamdorf and J. C. Hall, *Acquiring surgical skills,* British Journal of Surgery **87**, 28 (2000).

[14] C. A. Pellegrini, A. L. Warshaw, and H. T. Debas, *Residency training in surgery in the 21st century: A new paradigm,* Surgery **136**, 953 (2004).

[15] D. A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development* (Prentice Hall, 1983).

[16] P. T. Engels and C. de Gara, *Learning styles of medical students, general surgery residents, and general surgeons: implications for surgical education,* BMC Medical Education **10** (2010), 10.1186/1472-6920-10-51.

[17] S. M. B. I. Botden and J. J. Jakimowicz, *What is going on in augmented reality simulation in laparoscopic surgery?* Surgical Endoscopy **23**, 1693 (2008).

[18] S. Bann, A. Darzi, Y. Munz, B. D. Kumar, and K. Moorthy, *Laparoscopic virtual reality and box trainers: is one superior to the other?* Surgical Endoscopy **18**, 485 (2004).

[19] M. P. Schijven, J. J. Jakimowicz, I. A. M. J. Broeders, and L. N. L. Tseng, *The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum,* Surgical Endoscopy **19**, 1220 (2005).

[20] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, and R. M. Satava, *Virtual reality training improves operating room performance,* Annals of Surgery **236**, 458 (2002).

[21] A. S. Thijssen and M. P. Schijven, *Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees?* The American Journal of Surgery **199**, 529 (2010).

[22] A. Gallagher and R. Satava, *Virtual reality as a metric for the assessment of laparoscopic psychomotor skills,* Surgical Endoscopy **16**, 1746 (2002).

[23] A. Zhang, M. Hünerbein, Y. Dai, P. M. Schlag, and S. Beller, *Construct validity testing of a laparoscopic surgery simulator (lap mentor®),* Surgical Endoscopy **22**, 1440 (2007).

[24] M. P. Schijven, J. J. Jakimowicz, I. A. M. J. Broeders, and L. N. L. Tseng, *The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum,* Surgical Endoscopy **19**, 1220 (2005).

[25] J. J. Jakimowicz and S. Buzink, *Training curriculum in minimal access surgery,* in *Training in Minimal Access Surgery* (Springer London, 2015) pp. 15–34.

[26] S. Buzink, M. Soltes, J. Radonak, A. Fingerhut, G. Hanna, and J. Jakimowicz, *Laparoscopic surgical skills programme: preliminary evaluation of grade i level 1 courses by trainees,* Videosurgery and Other Miniinvasive Techniques **3**, 188 (2012).

[27] G. M. Fried, *FLS assessment of competency using simulated laparoscopic tasks,* Journal of Gastrointestinal Surgery **12**, 210 (2007).

[28] B. M. Garrett, M. MacPhee, and C. Jackson, *Implementing high-fidelity simulation in canada: Reflections on 3years of practice,* Nurse Education Today **31**, 671 (2011).

[29] K. D. Frick, S. M. Swoboda, K. Mansukhani, and P. R. Jeffries, *An economic model for clinical simulation in prelicensure nursing programs,* Journal of Nursing Regulation **5**, 9 (2014).

[30] A. Almeida, F. Rebelo, and P. Noriega, *In situ simulation: Challenges and results,* in *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 3: Performance and Tools)* (Agency for Healthcare Research and Quality (US), 2008).

[31] S. B. Issenberg and R. J. Scalese, *Simulation in health care education,* Perspectives in Biology and Medicine **51**, 31 (2007).

[32] D. Miskovic, M. Ni, S. M. Wyles, R. H. Kennedy, N. K. Francis, A. Parvaiz, C. Cunningham, T. A. Rockall, A. M. Gudgeon, M. G. Coleman, and G. B. Hanna, *Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England,* Annals of Surgery **257**, 476 (2013).

[33] M. C. Vassiliou, L. S. Feldman, C. G. Andrew, S. Bergman, K. Leffondré, D. Stanbridge, and G. M. Fried, *A global assessment tool for evaluation of intraoperative laparoscopic skills,* The American Journal of Surgery **190**, 107 (2005).

[34] S. Ganni, S. M. B. I. Botden, M. Chmarra, R. H. M. Goossens, and J. J. Jakimowicz, *A software-based tool for video motion tracking in the surgical skills assessment landscape,* Surgical Endoscopy **32**, 2994 (2018).

[35] V. N. Palter, N. Orzech, R. K. Reznick, and T. P. Grantcharov, *Validation of a structured training and assessment curriculum for technical skill acquisition in minimally invasive surgery,* Annals of Surgery **257**, 224 (2013).

[36] K. E. Roberts, *Evolution of surgical skills training,* World Journal of Gastroenterology **# 12**, 3219 (2006).

[37] A. G. Gallagher, E. M. Ritter, H. Champion, G. Higgins, M. P. Fried, G. Moses, C. D. Smith, and R. M. Satava, *Virtual reality simulation for the operating room,* Annals of Surgery **241**, 364 (2005).

[38] H. W. R. Schreuder, G. Oei, M. Maas, J. C. C. Borleffs, and M. P. Schijven, *Implementation of simulation in surgical practice: Minimally invasive surgery has taken the lead: The dutch experience,* Medical Teacher **33**, 105 (2011).

9

TRAINEE SATISFACTION WITH CHANGING CURRICULUM MODALITIES

9

Introduction

Trainee feedback is an important element in the curriculum design process. To understand the impact of a curriculum it is essential to survey the feedback in participant satisfaction where different strengths, weaknesses and perceptions of a curriculum can be uncovered. During the course of this research all the participant feedback was obtained in addition to the protocols of the research. This was considered a crucial element to monitor the change in trainee perceptions of the curriculum with the change in modalities implemented at different intervals over the period of this research.

Methods

The participant surveys were conducted as part of the routine LSS curriculum 2-day program from the period of 2013 to 2018 conducted in the Netherlands, India, France and Romania. A total of 173 responses were collected during the period. The change in modalities brought about by this PhD research were implemented from early 2015. The data was thus gathered to compare participant surveys from before the implementation of changes in curriculum to later. The feedback form was subdivided into three categories consisting of pre-course, course and post-course evaluation. The pre-course consisted of the trainee perception of the materials made available prior to the course, the time and guidance they received to prepare sufficiently for the program, the relevancy of the course materials to their current proficiency. The course evaluation consisted of their perception on course content, modules, modalities, pace of the course and assessment patterns. The post course evaluation reflected on the overall satisfaction and their perception of implanting the learning outcomes in practice.

SURVEY

The survey as administered is illustrated as follows:

T L

Course evaluation form on the LSS course Grade 1 Level 1

Course location: ☐

Date: ...... / ...... / 201......

Participant code (optional): ☐

A. Pre-course

1. **Rate on a scale from 1 (disagree) till 5 (agree) the following:**

	1 (Disagree)	2	3	4	5 (Agree)	N/A
The written material handed out before the course was easily accesible	☐	☐	☐	☐	☐	☐
I found the content of the written material relevant	☐	☐	☐	☐	☐	☐
The quality of the written material was satisfactory	☐	☐	☐	☐	☐	☐
The knowledge test reflected the written material	☐	☐	☐	☐	☐	☐
The questions in the knowledge test were relevant	☐	☐	☐	☐	☐	☐
Overall the quality of the knowledge test was satisfactory	☐	☐	☐	☐	☐	☐
The difficulty of the knowledge test was appropriate	☐	☐	☐	☐	☐	☐
The knowledge test motivated me to study the written material	☐	☐	☐	☐	☐	☐
I found the time spent to read the material and pass the test to be sufficient	☐	☐	☐	☐	☐	☐
I prefer to study theory before the course instead of listening to lectures during the course, to save one course day	☐	☐	☐	☐	☐	☐
The dexterity test at the start of the course motivated me to train my psychomotor skills before the course	☐	☐	☐	☐	☐	☐

B. The course

2. **Rate on a scale from 1 (disagree) till 5 (agree) the following:**

	1 (Disagree)	2	3	4	5 (Agree)	N/A
The course fitted to my expertise level	☐	☐	☐	☐	☐	☐
My expertise level matched that of the other participants	☐	☐	☐	☐	☐	☐
The topics were well chosen	☐	☐	☐	☐	☐	☐
It was well doable to go through all the course elements within the time specified	☐	☐	☐	☐	☐	☐
The balance between theory and hands-on training was good	☐	☐	☐	☐	☐	☐
The level of hands-on training matched the teaching material	☐	☐	☐	☐	☐	☐
The pace of the course was good	☐	☐	☐	☐	☐	☐
I expect to apply the course content in practice	☐	☐	☐	☐	☐	☐
The problem-based discussions were fruitful	☐	☐	☐	☐	☐	☐
The amount of practical training was good	☐	☐	☐	☐	☐	☐
The training exercises on the box trainers were relevant	☐	☐	☐	☐	☐	☐
The training exercises on the VR simulators were relevant	☐	☐	☐	☐	☐	☐
The difficulty of the simulator assessment was appropriate	☐	☐	☐	☐	☐	☐
The time for discussions were sufficient	☐	☐	☐	☐	☐	☐
The scenario-based assessment reflected the course content	☐	☐	☐	☐	☐	☐
The questions in the scenario-based assessment were relevant	☐	☐	☐	☐	☐	☐
Overall the quality of the scenario-based assessment was satisfactory	☐	☐	☐	☐	☐	☐
The difficulty of the scenario-based assessment was appropriate	☐	☐	☐	☐	☐	☐
I prefer a two days course instead of a three days course	☐	☐	☐	☐	☐	☐

L T

T L

3. **What features do you think are important in a basic skills course on a scale from 1**(not important) **till 5**(important)

	1 (not important)	2	3	4	5 (most important)	N/A
Practical training on box trainers with mock-ups	☐	☐	☐	☐	☐	☐
Practical training on VR simulators	☐	☐	☐	☐	☐	☐
Practical training on box trainers with animal tissue	☐	☐	☐	☐	☐	☐
Practical training on live animals	☐	☐	☐	☐	☐	☐
Practical training in general	☐	☐	☐	☐	☐	☐
Theoretical lectures	☐	☐	☐	☐	☐	☐
Problem-based discussions	☐	☐	☐	☐	☐	☐
Assessment of basic psychomotor skills	☐	☐	☐	☐	☐	☐
Assessment of hands-on procedural skills	☐	☐	☐	☐	☐	☐
Assessment of theoretical knowledge	☐	☐	☐	☐	☐	☐
Assessment of cognitive skills (judgement, team-behavior)	☐	☐	☐	☐	☐	☐
Other (please specify): ____________________:	☐	☐	☐	☐	☐	☐

C. After the course

4. **Rate on a scale from 1 (disagree) till 5 (agree) the following:**

	1 (Disagree)	2	3	4	5 (Agree)	N/A
I believe that a short time between the course and training in the clinic is beneficial	☐	☐	☐	☐	☐	☐
I think it will be feasible to send in two videotapes within 3 months after the course	☐	☐	☐	☐	☐	☐
The close follow-up by my supervisor after the course will be beneficial to my time as a resident	☐	☐	☐	☐	☐	☐

D. Overall

5. **Rate on a scale from 1(disagree) till 5(agree) the following:**

	1 (Disagree)	2	3	4	5 (Agree)	N/A
The course location was suitable for the course	☐	☐	☐	☐	☐	☐
The number of breaks were sufficient	☐	☐	☐	☐	☐	☐
The level of expertise of the faculty was sufficient	☐	☐	☐	☐	☐	☐
The faculty and the staff took good notice of my questions and difficulties	☐	☐	☐	☐	☐	☐

6. **What features do you think are important in a basic skills course on a scale from 1 till 5** (*1 is not important and 5 is most important*).

	1 (not important)	2	3	4	5 (most important)	N/A
An efficient course	☐	☐	☐	☐	☐	☐
Accreditation by the european association for endoscopic surgery (EAES)	☐	☐	☐	☐	☐	☐
The intention of being allowed to operate shortly after the course and approval of videos	☐	☐	☐	☐	☐	☐
Assuring qualified surgeons by assessment of skills	☐	☐	☐	☐	☐	☐
Other (please specify): ____________________:	☐	☐	☐	☐	☐	☐

7. **What would you add to the course if you were to add something?**

L T

T L

8. What would you remove from the course if you were to remove something?

If you have any additional comments, please add them here:

Thank you for your feedback!

L

T

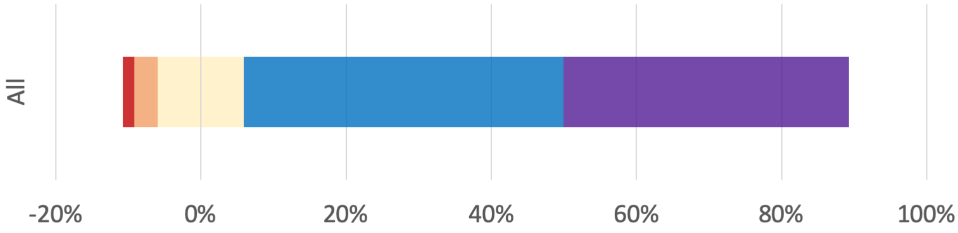


Figure 9.1: Summary of the overall data with Likert colorings from 1=red to 5=purple

STATISTICS

The hand-completed data were inputted into Matlab (R16b) where they were analysed and presented as follows.

RESULTS

PARTICIPANTS

In total, responses from 173 participants were recorded across a period of 6 years in four locations: Cluj, Eindhoven, Lyon and India. Between the 2015 and 2016 surveys, improvements were made to the delivery of the course based on the feedback; the data can thus be considered in two groups: i) before and ii) after the changes – in which there were 92 and 82 participants recorded, respectively.

OVERALL SATISFACTION

Considering all questions, the overall level of satisfaction was very high as can be seen in Figure 9.1, more than 82.97% of the responses were ranked as a 4 or 5. The average response was 4.16.

9

Considering the before and after groups separately, we see from Figure 9.2 that there is a significant increase in the average score, as well as a change in the skew of the distribution (towards greater satisfaction, especially the number of 5s increases).

INDIVIDUAL QUESTIONS

Looking at the questions individually we see significantly ($p<0.05$) increased average in 13 cases, specifically questions 21, 22, 23, 24, 26, 27, 28, 29, 38, 42, 44, 46 and 48. No results were significantly decreased on average after the changes.

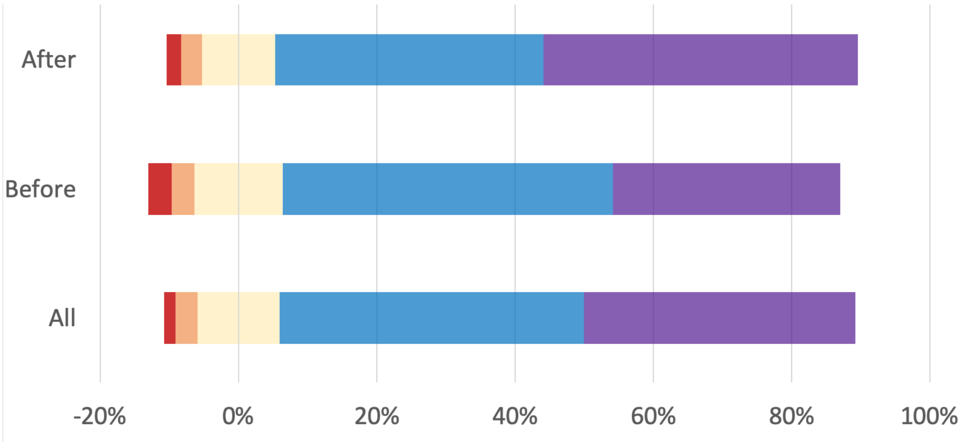
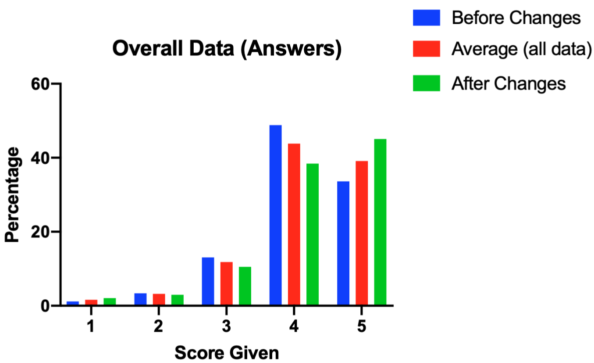
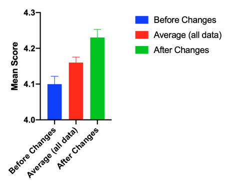


Figure 9.2: Summary of the overall data and that data grouped into those responses before and after the changes with Likert colorings from 1=red to 5=purple



(a) Distribution of responses to all questions and grouped by when the changes were made.



(b) Average response with SEM (each bar is statistically different from the others ($p<0.05$))

Figure 9.3: Overall data categorised by response

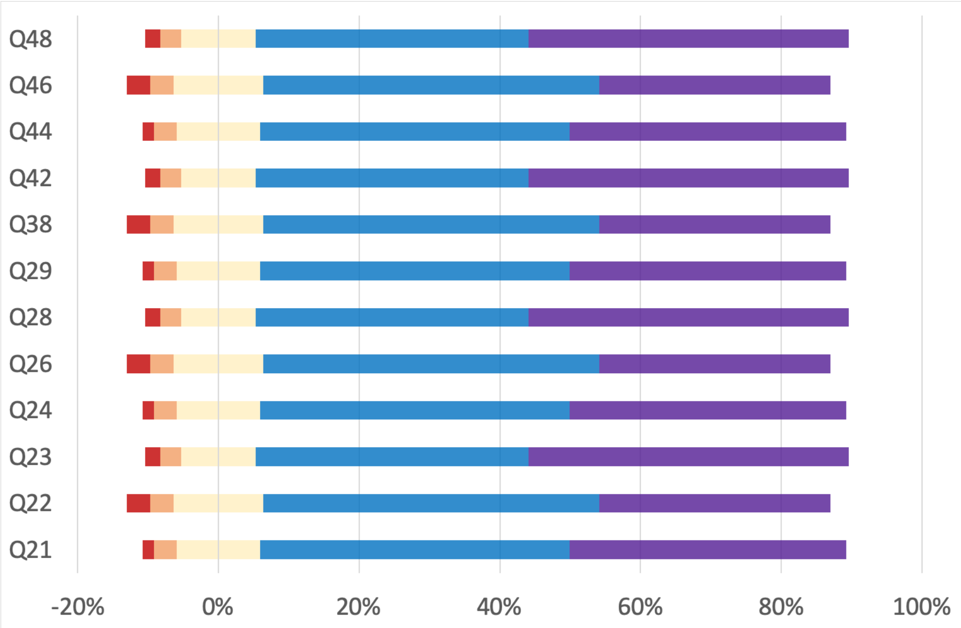


Figure 9.4: Summary of the thirteen 'improved' questions from after the changes with Likert colorings from 1=red to 5=purple

DISCUSSION

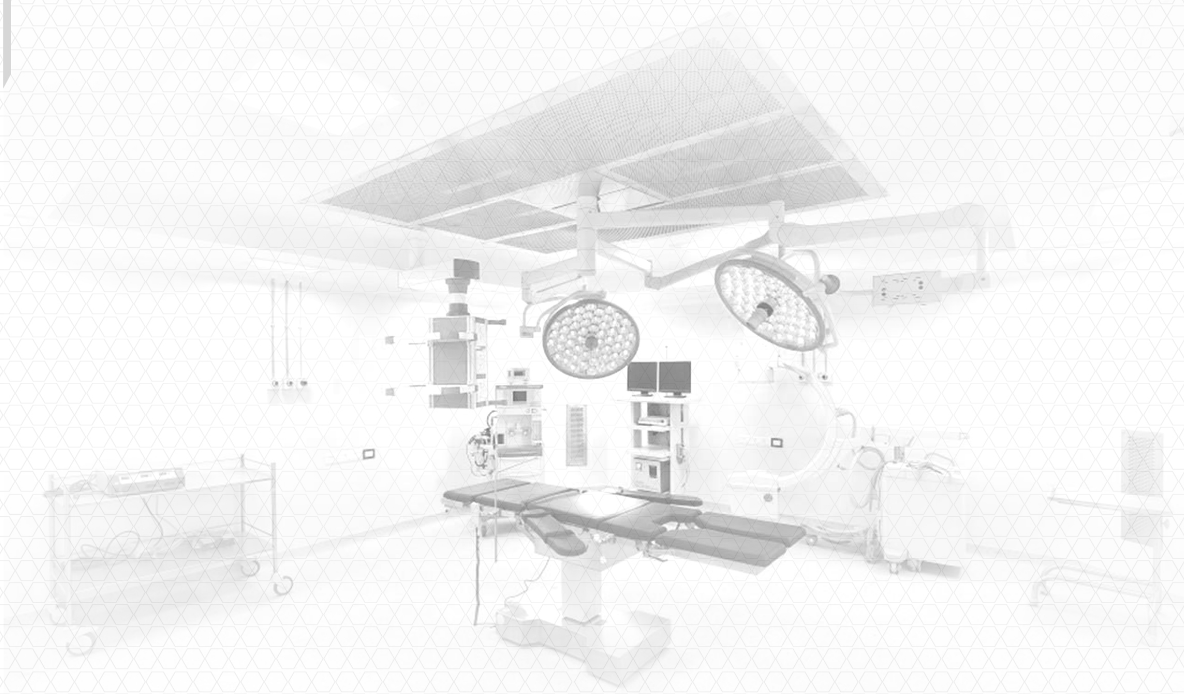
The aim of this study was to observe whether changes made to the LSS course Grade 1 Level 1 had a positive impact on users self-assessed feedback. The resulting data was anlaysed against the specific changes made in the curriculum design and implementation. Further, these change in modalities were a result of participant feedback and the outcomes of research based on the LSS curriculum. The results showed that participants perceived the course to be better adapted to their needs and learning speed showing the benefit of consistent and evidence-based improvement.

9

Overall level of satisfaction for the course was good before the adjustments, at nearly 83%. When we compare the data grouped by whether their surveys were submitted before (Figure 9.2 and 9.3) or after the changes, we see that there is a significant rightward shift in the after results (from 82% to 85%), suggesting individuals self-assess more positively after the changes than before.

Looking at the individual questions, once grouped, we see statistically significant improvements in 13 of the questions and no questions were significantly worse, further suggesting that, overall, the changes were positive.

10



EFFECTS OF SURGICAL FLOW DISRUPTIONS ON THE SURGEONS' RESOURCES: A PILOT STUDY

B.C.G. van Houwelingen, A.F.Rutkowski, S.Ganni, P.S.Stepaniak, J.J.Jakimowicz

Published in Surgical Endoscopy, November, 2019

10

ABSTRACT

Background Minimally Invasive Surgery (MIS) requires surgeons to allocate more attention and efforts than open-surgery. A surgeon's pool of resource is affected by the multiple occurrences of interruptions and distractions in the Operating Room (OR). Surgical flow disruption has been addressed from a quantitative perspective. However, little is known on its impact on the surgeons' physiological resources.

Methods Three physiological markers Heat Flux (HF), energy expenditure in Metabolic Equivalent of Tasks (METs) and Galvanic Skin Response (GSR) were recorded using body sensor monitoring during the 21 surgical operation. The three markers respectively represent: stress, energy mobilization and task engagement. A total of 8 surgeons with different levels of expertise (expert vs. novice) were observed performing 21 surgical procedures categorized as short versus long. Factors of distractions were time-stamped, and triangulated with physiological markers, Two cases illustrate the impact of surgical flow disruptions on the surgeons.

Results The results indicate that expert surgeons' mental schemata are better organized than novices. Additionally, the physiological markers indicate that novice surgeons display a higher HF at the start (tendency p=.059) and at the end of procedures (p=.001) when compared to experts. However, during longer procedures

expert surgeons have higher HF at the start (p=.041) and at the end (p=.026), than at the start and end of a short procedures.

Conclusion Data collected during this pilot study showed that interruptions and disruptions affect novice and expert surgeons differently. Surgical flow disruption appears to be taxing on the surgeons' mental, emotional and physiological resources; as a function of the length and nature of the disruptions. Several training curricula have incorporated the use of virtual reality programs to train surgeons to cope with the new technology and equipment. We recommend integrating interruptions and distractions in virtual reality training programs as these impact the surgeons' pool of resources.

INTRODUCTION

Classical literature in the field of surgery demonstrates that Minimally Invasive Surgery (MIS) is more demanding on the surgeons' resource than open-surgery [1–4]. Surgery is a stressful profession [5]. MIS enhance treatment capabilities, placing, however, an ever-increasing pressure on the surgeons [6, 7]. Resources are an individually-possessed form of physical, emotional or cognitive energy required in processing information [8, 9]. Resources are limited; thus, they form a pool and affect each other through a feedback loop [10, 11]. The complexity of technologies in the OR requires surgeons to allocate their resources mindfully to reach optimal surgical results [3]. Particularly, overloaded surgeons may lose their abilities to maintain patient safety in the Operating Room [12].

Distractions and interruptions are an additional well-known burden on the surgical performance. For example, Wiegmann et al. [13] demonstrated that flow disruptions such as teamwork/communication failures, equipment and technology problems, extraneous interruptions, and training-related distractions led to surgical errors. Environmental factors (e.g., equipment design), social factors (e.g., teamwork, communications), and organizational factors (e.g., scheduling, procedures and policies) are as much potential distractors [12, 14–18]. Disruptions in the OR have mostly been studied from a quantitative perspective in relationship to surgical errors. For example, Zeng et al. [12] studied the frequency and duration of disruptive events (e.g., instrument change, surgeon position change, extraneous interruption) on surgical delay. Using video-aided observations the authors demonstrated that on average, disruptive events performed in the OR caused 4.1 min of delay for each case per hour, corresponding to 6.5% of the procedure time: instrument change (3.4 min/h) generate the most surgical delays. In a recent article Al-Hakim et al. [19] used a similar approach, focusing on the impact of ergonomics factors (e.g., monitor location, level of instruments' handles, and location of surgical team members) on the operative flow disruption. The literature also reports that paying attention and responding to alarm increases the surgeons' mental load and stress level [20]. It creates a competition for attentional resources. These multiple factors disrupt the natural progression of an operation, potentially compromising patient safety [13, 21]. The recurrent disruptions of the surgical flow lead to an increase in surgical errors and impact surgeons' mental strain [22, 23]. Understandably, the more disruptions the more the surgeon must tap into his pool of resources to alleviate potential negative effects. However, surgeons' experience of interruptions and distractions differ in practice. For example, noise is a recognized source of stress, and impairs concentration and communication in the OR. Still, some surgeons may enjoy music in the theater while others require a quiet environment [24].

Surgeons are able to recognize most disturbing factors, but have a hard time quantifying or sequencing these factors objectively. That is in practice, surgeons report various levels of resistance to disruptions when engaged on the topic. They may experience objective (i.e., physiological level) repetitive stress without consciously identifying it at the subjective level (i.e., verbal report). They are "implicitly" trained in coping emotionally and cognitively with these physiological modifications. Congruently, surgeons do not systematically perceive all distractions and interruptions as consuming their attention. Additionally, they often fail to recognize that they suffer from stress [25, 26]. However, research has demonstrated that excessive and long-lasting stress compromise the surgeons' technical and non-technical skills (e.g., teamwork, decision-making) [27, 28]. Weenk et al. [26] used wearable sensors to collect the Heart Rate Variability (HRV) of surgeons. They concluded that the stress was highest performing an operation in fellows and residents than in consultants. Interestingly, Weenk et al, [26] results showed that the self-reported stress level (i.e., State Trait Anxiety Inventory) did not correlate with the physiological measurements (i.e., HRV).

In this article, we assume that surgeons may not be fully aware of the impact of disruptions and interruptions in the OR on their stress level, and therefore, on the surgical performance. The pilot study presented in this paper focuses on the impact of surgical flow disruptions on the surgeons' cognitive, emotional and physiological resources. We speculate that surgeons who possess high level of expertise and skills (i.e., cognitive resources), nerves of steel when for example dealing with severe bleeding (e.g., emotional resources), after long hours of surgical procedure (i.e., physiological resources) will see their pool of resources particularly challenged when they have to cope with repetitive disruptions of the surgical flow.

	Expert surgeons (from 5 to 19 years in function)	Novice surgeon in training (from study year 2–5 to 6 months of experience)
Short procedure < 1 h ($n = 14$)	Gastric bypass (2) Gastric sleeve (1) Hernia (1) Diagnostic laparoscopy (2)	Gastric sleeve (1) Hernia (1) Cholecystectomy (6)
Long procedure > 4 h ($n = 7$)	Esophagus (5) Tumor resection (1)	Lap Nissen fundoplication (1)

Table 10.1: Types of procedures observed and the level of experience of the surgeons in years of practice

MATERIALS AND METHODS

The research was conducted in the department of Surgery and department of research and education at Catharina hospital in Eindhoven (The Netherlands) over a period of six months. The surgeons and members of the surgical team were informed of the goal of the research. Consents were collected prior to the procedure. The pilot study was approved by the ethics committee of the hospital.

PROCEDURE

Disruptions, interruptions of the surgical flow as well as surgeons' physiological markers of stress, energy mobilization and task engagement were collected using the SenseWear Pro 3 armband during twenty-one surgical procedures representing approximatively 21 hours of observation. The surgical flow disruptions reported could be later triangulated with the measurements gathered with the physiological measurements collected with the SenseWear Pro 3 armband. The surgeon was equipped with the wearable prior to scrubbing and going sterile. Physical activities (e.g., stretching, yawning, laughing, walking, pulling or pushing of tissue or the patient) was consigned in the observation file in addition to the observed distractions and interruptions. Gender, age, weight, handedness and smoking were recorded for reliability purpose. Following the observations, the surgeon was invited for a short debriefing with the observer. Information regarding the surgical procedures such as the type of surgical procedure, the start of procedure (i.e., time of the first incision), the end of the procedure (i.e., start of the final stitching) as well as the team composition (i.e., members and roles) were consigned. Table 10.1 presents the type of procedures and the level of experience of the surgeons in years of practice.

MEASUREMENTS

A non-exhaustive list of disruptions and interruptions was built based on the literature to report the observations during the 21 surgical procedures [12, 14–18]. The list of factors has been pre-tested with the participation of four expert surgeons, and later complemented by a set of observations conducted in the operating room. Each real-time occurrences of disruptions were reported and time-stamped in an excel file. The list was composed of environmental factors (e.g., operating room environment, environmental hazards), social factors (e.g., teamwork, communications), equipment factors (e.g., technologies and instruments, technical default), organizational factors, training and knowledge factors (e.g., technical factors, training and procedures).The SenseWear provided three physiological markers measurements: (i) the Heat Flux (HF) that is the amount of heat that is being dissipated from the body via the skin [29]. The Heat Flux is classically used an indicator of stress. The HF scale range in from 0.00 W/m2 to 300W/m2. A two-standard-deviation range of +/-10.00W/m2 at HF inferior than 50W/m2. In this pilot study the Heat Flux ranged between 40W/m2 and 110W/m2. Previous research has demonstrated that difficulties during surgeries, e.g. distractions, influence the stress level increasing the Heat Flux level [25, 26, 30].; (ii) the METs value that is the physiological measures of energy expenditure in Metabolic Equivalent of Tasks. The METS scale range between 56KW to 20MW. A two-standard-deviation range of +/- 3.00% of expected value. This measure allows controlling the influence of physical activity on galvanic skin response [29–31]; (iii) the Galvanic Skin Response (GSR) that is the electrical conductivity of the skin. The GSR score range from 20ºC to 40º. A 2 standard deviation range of +/- 0.80°C across the temperature range. Skin conductance level is a reliable indicator for the level task engagement [29]. Increase in task complexity relates to more task engagement [32]. Distractions in the OR increase the complexity of the procedure as it increases the cognitive resources needed to complete the task, potentially increasing the GSR. The GSR spikes allow assessing task engagement at certain point of the procedure. It is related to the METS value providing a good indicator on energy expenditure.

RESULTS AND DISCUSSION

SURGICAL FLOW DISRUPTION: OCCURRENCE OF DISTRACTION AND INTERRUPTION EVENTS

A total of 1541 distracting events were recorded during the 20 hours 19 minutes and exactly 06 seconds of observation. The three top distractions computed through the 21 surgical procedures were instruments change (30.7%); procedure or patient irrelevant communication (13, 9%); Operating Room door opening (12.8%). Radio conversation, phone communication as well as sounds of alarm represented all together another set of disrupting factors (16%). Most have been reported in the literature [12, 14–18].

PHYSIOLOGICAL MARKERS: HEAT FLUX, METS AND GSR

The three physiological markers were collected continuously through the whole duration of the procedure for each of the surgeons. As previously reported, Heat Flux is an indicator of stress; this measurement may serve as a proxy of the emotional resources required as part of the surgeon profession [26]. METs is related to physical activation; this measurement is an interesting indicator to assess the physiological resources required to perform the surgical tasks. The GSR indicates modifications in task engagement; this measurement concerns mostly the cognitive resources required to deploy efficiently (e.g., effortless) the cognitive schemata required for the surgery. These three measurements combined indicate how the surgeon's body consumes fuel to cope with complex tasks and situations in the OR. The data collected allowed comparisons between expert and novice surgeons engaged in short procedure, as well as long procedure.

A conservative statistical approach was used to analyze the continuous outcome of the three physiological variables [33]. Non-parametric tests equivalent of parametric tests was selected as the appropriated statistical tools regarding the small size sample, and despite the continuous outcome of the three physiological measurements. Indeed, due to the reliance on fewer assumptions, non-parametric methods are more robust [34]. The Mann-Whitney U test is the non-parametric equivalent of the two sample T-test; the Wilcoxon signed rank test of the paired T-test and the Spearman's rho is the equivalent of the Pearson correlation. Tables 10.2, 10.3, 10.4 and 10.5 present the median values as well as the minimum and maximum values for each of the three markers. Additionally, we present for the sake of readability the mean scores and the standard deviation. Indeed, these parametric values are informative and reliable (e.g., computation of continuous outcome).

First, we compared the average scores of the three physiological markers as a function of the levels of experience during the short procedures. Indeed, in this pilot, we add no situation of novice surgeons involved in long surgical procedure. Second, we compared the average scores of the three physiological markers within each of the three conditions observed. The results of the Mann-Whitney U test (equivalent two sample T-test) show that novices display a higher Heat Flux at the start of the procedures (tendency p=.059) and at the end of the procedures (p=.001) than the expert. Table 10.2

	HF at start of procedure*	HF at end of procedure**
Novice***		
N	8	8
Mean	84.8468	92.6925
Median	87.2830	90.5656
SD	7.62125	8.19507
Min	70.67	82.52
Max	94.03	104.53
Expert ns.		
N	6	6
Mean	70.2812	59.2122
Median	69.7486	57.3030
SD	14.56966	12.43005
Min	53.95	43.95
Max	93.83	76.11

*Ns $p = .059$; **p. < .001; ***$p = .046$; ns. $p = .091$

Table 10.2: Heat Flux (i.e., stress) at the start and at the end of the short procedure: Expert vs novice

presents the results of the mean, median, minimum, and maximum value of the surgeons' HF at the start and at the end of the short procedure: Novice versus Expert.

The results of the Mann-Whitney U test did not show significant differences between novice and expert surgeons in METs and GSR. The energy mobilized as well as the overall task engagement appeared to be similar for expert and novice surgeons. The results of the Wilcoxon signed rank test (equivalent of paired T-test) indicate that the Heat Flux of the novice surgeons increased significantly between the start and end of the procedure (p=.046) while it slightly but not significantly decreased for the expert (p=.091). The results indicate that the novice surgeons experience more stress than experts did at the start, anticipating the surgical procedure, as well as during the whole procedure. Also, the results indicate that the novice surgeons experience more stress than experts did at the start, anticipating the surgical procedure, as well as during the whole procedure. This result is congruent with previous research using HVR as a proxy of stress [26].

Interestingly, the results of the Wilcoxon signed rank test indicate that the METs remained at constant value for expert surgeons (p=.753), while their GSR increased significantly for short (p=.028) and long (p=.0281) surgical operation. The same analysis conducted for the group of novice surgeons revealed that METs score also decreased slightly but not significantly (p=.09), while the GSR increased drastically and significantly (p=.012). Table 10.3 presents an overview of the results of the mean, median, minimum, and maximum value of the surgeons' GSR at the start and at the end of the procedure.

	GSR at start of procedure	GSR at end of procedure
Short novice*		
N	8	8
Mean	0.0767	0.1703
Median	0.0755	0.1701
SD	0.02724	0.05841
Min	0.05	0.11
Max	0.12	0.26
Short expert**		
N	6	6
Mean	0.0567	0.1267
Median	0.0436	0.1301
SD	0.03363	0.04907
Min	0.02	0.07
Max	0.11	0.20
Long expert***		
N	6	6
Mean	0.0401	0.1330
Median	0.0366	0.1297
SD	0.03006	0.02703
Min	0.01	0.10
Max	0.07	0.18

*$p = .012$; **$p = .028$; ***p. = .0281

Table 10.3: GSR at the start and at the end of the procedure for the three categories of observations

The energy mobilization measured through METS remained mostly constant for both novice and expert surgeons. This result is not surprising as surgeons operate in a static position. Table 10.4 presents the values for the METS. NO significant effects have been found for this physiological marker.

It appears clearly that the METS values are to be interpreted in light of the GSR. The GSR is a reliable indicator for the level task engagement increased significantly for both groups, and this more drastically for the novice surgeons. These results confirm the assumption according to the levels of expertise impact the amount of stress and task engagement to perform short MIS surgical procedure. Finally, we compared the average scores of the three physiological markers as function of the length of the procedure: short vs. long for the group of expert surgeons. The results of the Mann-Whitney U test indicate that the expert surgeons operating on a long procedure have a significantly higher Heat Flux at the start (p=.041) and at the end (p=.026), than they do at the start and end of a short surgical procedure. Table 10.5. presents the results for the expert surgeons Heat Flux values for both short and long procedures.

The results indicate that stress by anticipation of expert surgeons is higher for long surgical procedure than in short procedure, and that surgeons end up more stressed at the end of a long procedure than a short one. This result is as previously underlined

	METs at start of procedure	METs at end of procedure
Short novice		
N	8	8
Mean	1.5137	1.3514
Median	1.5692	1.2729
SD	0.25100	0.28166
Min	1.01	1.05
Max	1.74	1.74
Short expert		
N	6	6
Mean	1.6267	1.6124
Median	1.6811	1.6164
SD	0.23149	0.13156
Min	1.22	1.45
Max	1.89	1.80
Long expert		
N	6	6
Mean	1.9748	1.5516
Median	1.7243	1.6155
SD	0.49686	0.20094
Min	1.55	1.15
Max	2.73	1.68

Table 10.4: METs at the start and at the end of the procedures

	HF at start of procedure*	HF at end of procedure**
Short expert		
N	6	6
Mean	70.2812	59.2122
Median	69.7486	57.3030
SD	14.56966	12.43005
Min	53.95	43.95
Max	93.83	76.11
Long expert		
N	6	6
Mean	87.7887	80.3255
Median	87.5897	79.8751
SD	8.28090	13.22641
Min	79.19	60.87
Max	100.30	98.57

*p = .041; **p = .026

Table 10.5: Expert surgeons Heat Flux values at the start and at the end of the procedure: Short vs long

10

congruent with the conclusion of Weenk et al [26] The results of the Wilcoxon signed rank test indicate that the Heat Flux of the surgeons remain stable through the procedure (p=.345). However, they start from a higher level than they do when engage in a short procedure.

The METs slightly decreased but not significantly, while the GSR significantly increase (p=.028) as it does in the case of a shorter procedure (see Table 10.3). To conclude, the stress level and level of task engagement are affected differently for short or long procedure. These results demonstrate that long surgical procedure impact negatively the amount of stress per anticipation. However, regardless of the length of the surgical procedure, task engagement seemingly increases during the course of the procedure.

GSR, METs, Heat Flux (HF) and distracting events occurrences

Two specific cases illustrate the impact of surgical flow disruptions on the surgeons' physiological markers; as a function of the length of the operation. We selected these two cases to illustrate from a qualitative rather than quantitative perspective the impact of surgical flow disruption on the surgeons' emotional, cognitive and physiological resources. A coding application allowed reporting with a real time-stamp the occurrence of the interruptions and distracting events during the 21 surgical procedures. The real time lapse occurring between the start and end-time of a distracting event served to build sets of instantiations of the surgical flow disruptions. In the next section we present examples of such instantiations triangulated with the three physiological measurements.

In order to better understand the impact of the distractions on the surgeons' task engagement and stress level, we combined the GSR measurements as well as the Heat Flux with the occurrence of distractions. Interestingly, one can imagine that distracting events are not to be observed under a sequential form. That is often they are observed under multiple, recurring and parallel occurrences as demonstrated in the case of these two surgeons. We purposely selected two representative cases to enlighten the results of the data collected for the overall sample.

Surgeon A Experienced (6 years as a surgeon), surgical procedure (gastric bypass), length of the operation (59 mn), total amount of distraction (91). The energy mobilization and stress level of the surgeon are significantly and positively related. Indeed, the results of the Spearman's rho (equivalent Pearson correlation) indicated that METs and the Heat Flux (HF) are positively and significantly correlated (r=.282,

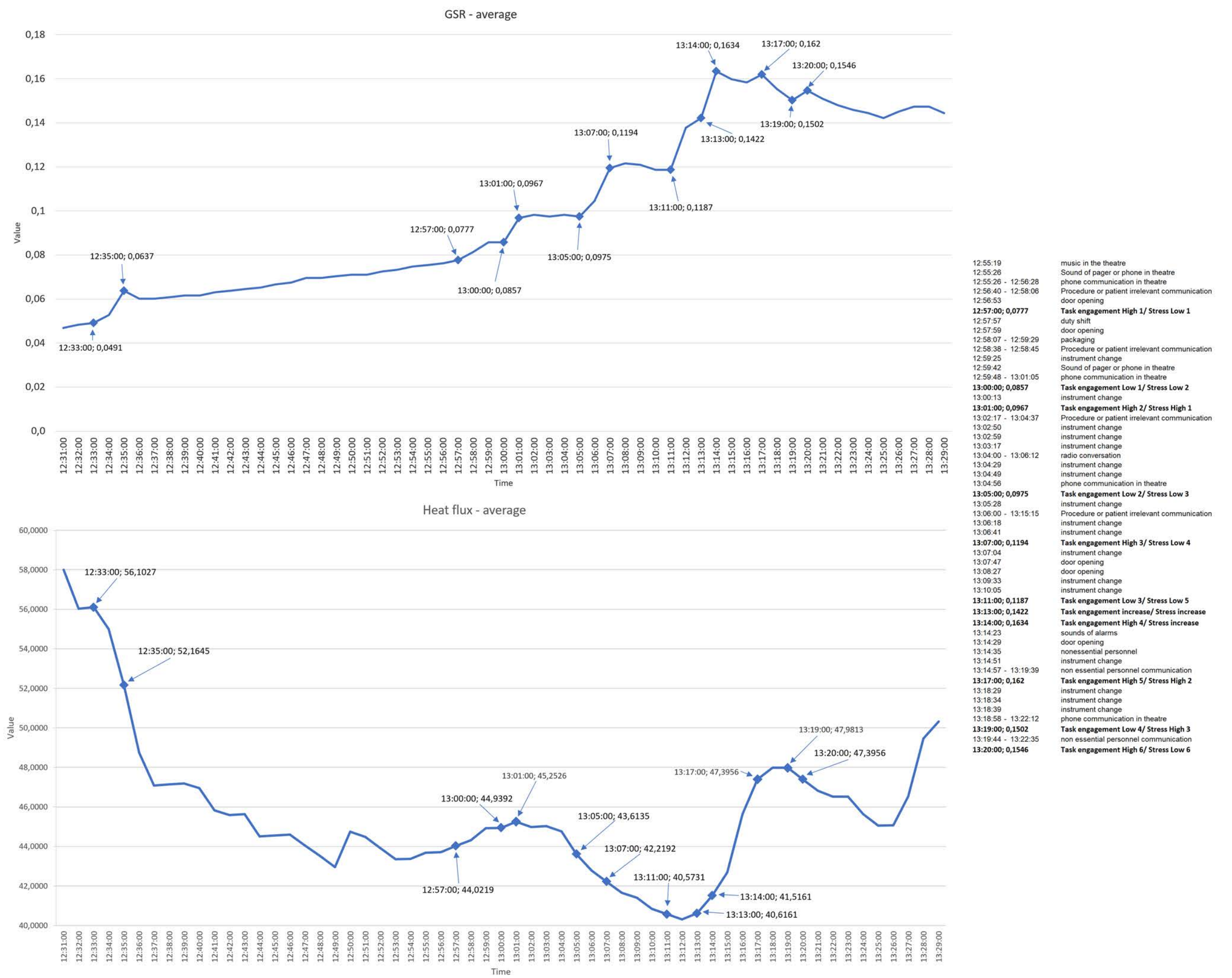


Figure 10.1: Surgeon A: GSR, Heat Flux (HF) and distracting events observed in relation to highs and lows in GSR and HF

10

p=.013). That is the surgeon deployed energy coping with stress. Interestingly, his level of task engagement was inversely related to his level of stress while performing this short surgery and not related to energy mobilization. Indeed, the GSR is negatively and significantly correlated to the HF (r= -.476, p=.0001) and not significantly correlated to the METs (r= .047). We concluded that while stress impacted energy mobilization during this short procedure, task engagement did not. That is when the surgeon had to pull on his physiological pool of resources it was to cope with stress. He did not have to pull on extra physiological resources attending to the surgical task. This can be mainly explained as a result of his level of expertise. Interestingly, the engagement in the task, was negatively related to the less stress he experienced. Figure 10.1 presents the highs and lows in GSR (task engagement) and Heat Flux (stress) of surgeon A. Between the time-stamp 12:57 and 13:20 that correspond to the highest frequency of highs and lows observed, a total 35 distracting events were recorded.

The patterns of the task engagement (GSR) and stress (HF) results associated to the distractions indicate that the highs observed are mostly related to the changes of instruments, packaging, as well as communication mostly irrelevant to the patient. Interestingly, 3 of the highs in task engagement are observed in combination with lows, or decreased in stress level. The high level of stress at the beginning may be explained by the fact the expert surgeon does not know what will happen during the procedure (i.e., which difficulties may be encountered). After starting the procedure, the surgeon experienced a form of control of the situation. The surgeon then switched to high focus on the surgical performance and therefore the high level of stress decreased.

As presented in Figure 10.1b in two points of time was the engagement of the surgeon highs as well as his stress level. These points in time correspond to a set of changes of instruments that may have indicated an important point in the procedure. Overall for these patterns of data it seems that the conversation has mostly the role of decreasing stress at the time of highs in task engagement. Additionally, experimental studies have shown that stress level can be judged based on the analysis of GSR and speech signals. However, and more probable than not, the state of hyper focus of the expert surgeon resulted in a delayed effect visible after the resolution of the problem. Then conversations took place as a form of outlet of stress.

Surgeon B Experienced (7 years as a surgeon), surgical procedure (esophagectomy), length of the operation (96 mn), total amount of distraction (128). The level of energy mobilization and stress were not significantly related for surgeon B. The results of the Spearman's rho indicated that the METs and the Heat Flux (HF) are negatively but not

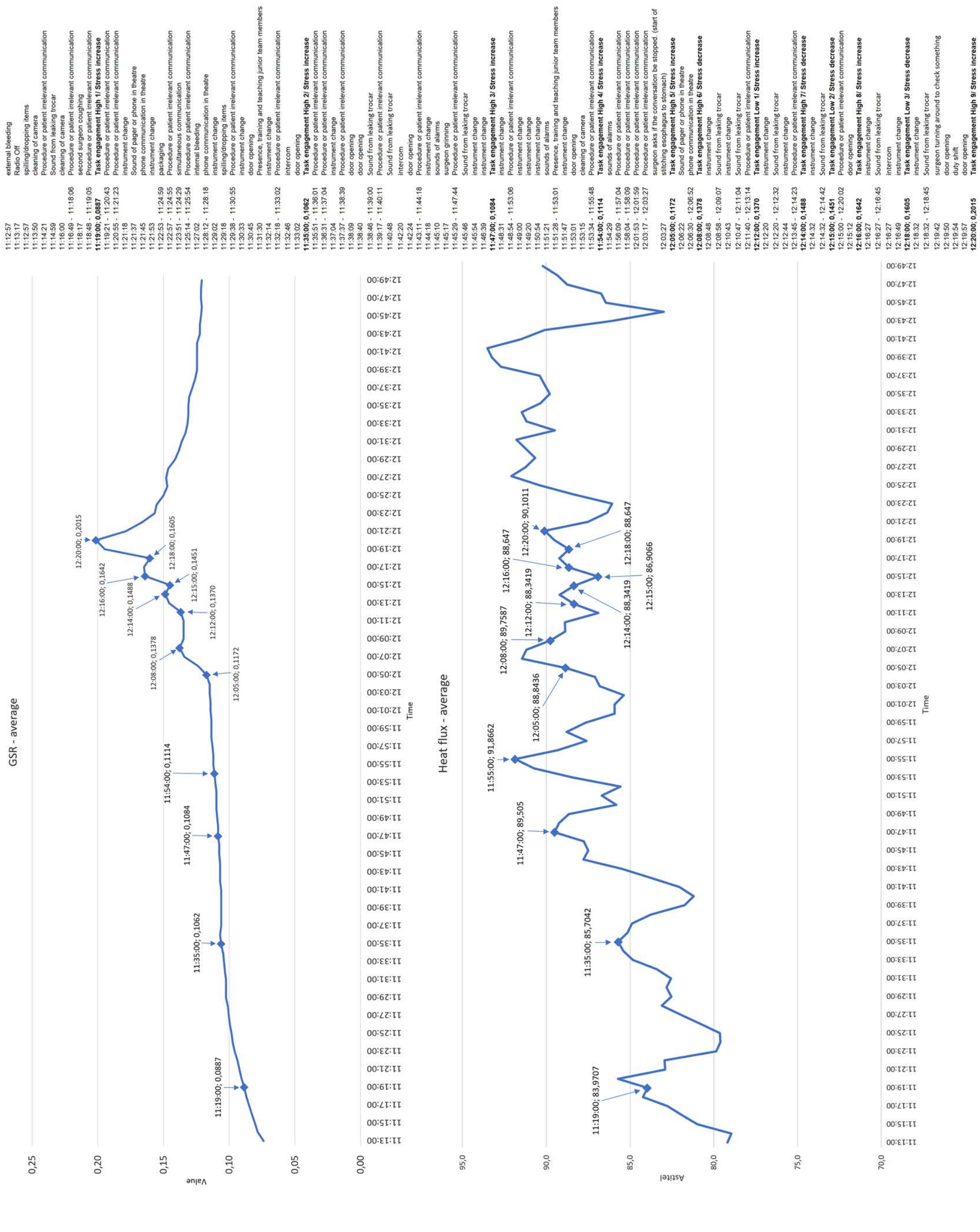


Figure 10.2: Surgeon B: GSR, Heat Flux (HF) and distracting events observed in relation to highs and lows in GSR and HF

significantly correlated (r= -.042, ns). In the case of surgeon B, it seems he did not have to pull on extra physiological resources to cope with the stress of a long operation. However, his cognitive engagement with the task was clearly related to his stress level. That is contrary to the case of surgeon A. Indeed, the results indicated that the HF is significantly correlated to the GSR (r=.613, p=.0001). The level of task engagement is negatively related to energy mobilization, however marginally. Indeed, the results show that the GSR is negatively correlated (tendency) to the METs (r= -.171, p=.072). As observed in the case of the surgeon A, surgeon B also did not have to pull on extra physiological resources attending to the long surgical task. This is supposedly related to the level of expertise. Striking is that during this long procedure, the more (or less) cognitively engaged the surgeon B was the more (or less) stress he experienced. Figure 10.2. presents the highs and lows in GSR and Heat Flux of an experience surgeon B performing a long surgery. Between the time-stamp 12:05 and 12:20 that correspond to the highest frequency of highs and lows observed in GSR, a total 26 distracting events were recorded.

As depicted in the Figure 10.2a, 4 major highs are observable in Heat Flux during this procedure that represent a form of accumulation of about 63 distractions. During that period, the GSR level kept rising steadily. These highs in Heat Flux could be mostly related to set of external bleeding, spilling/dropping items, procedure irrelevant communications, intercom, cleaning of the camera, trocar leakage, as well as sound of alarms. The patterns of observations are different for the case of surgeons A and B. Indeed, the association in term of the combination between highs and lows in GSR and Heat Flux are different. However, we can find similarities and point at specifics and interesting events. For example, the observations inform us that the highs 1 and 2 in task engagement occurred following the surgeon's request to keep quiet in the OR while he was anticipating a difficult point in the surgery (i.e., start of stitching esophagus to stomach). Also, the results indicate that the task engagement of surgeon B is indeed at a high point when the communication is meaningful to the procedure, and stress increased with unusual and irritating sounds such as the sound of the trocar. As in the case of surgeon A, decreased in stress is observable in association to irrelevant communication even when the surgeon is cognitively engaged. Moreover, we assumed as in Case A, a delay effect in GSR visible after the resolution of the problem, translating into communicative behaviors. Disruptions such as multiple door openings or duty shift led to an increase in the surgeon level of stress. The amount of distractions in relation to HF presented in Figure 10.2 demonstrated the clear negative influence of distractions on the surgeon's HF, even when he managed to remain highly cognitively engaged. The case of surgeon B indicated clearly that the task engagement was related to the fluctuation in the level of stress. The impact of combined distractions such as leaking trocar, door opening, or duty shift may have been a burden during the operation. The surgeon mentioned during the debriefing that the sound of the leaking trocar was annoying. As in the case of the surgeon B, irrelevant communication may have served as an outlet of stress, or indicating the end of difficult procedure point requiring high focus

Discussion and limitations

The peripheral nervous system regulates homeostatic processes such as body temperature and blood flow. Potential threats of our bodily homeostasis generate stress [35]. The three physiological markers address different psychological phenomena. While we could well relate Heat Flux to the level of stress, GSR to task engagement, the METs measurement appeared less informative in this research but for Novice surgeon. Also, the stress level and level of task engagement are affected differently for short or long procedure. Long surgical procedure impact negatively the amount of stress per anticipation. However, regardless of the length of the surgical procedure, task engagement seemingly increases during the procedure.

Training in a skills laboratory is key in acquiring intellectual and technical surgical skills require to perform surgery but also in preventing exhaustion while dealing with interruption of the surgical flow [36, 37]. Research demonstrated that when there is a lack of training, schemata automation is poorer, leading to a higher risk of failure and increasing the stress level of surgeons [30]. When task complexity increases trainees use more of their attentional resources concentrating on technical aspects of the task performance rather than on higher level activities (e.g., anticipating, scanning, or attending to instrument read-outs) [38]. The results of the research congruently demonstrated novice surgeons experienced more stress than experts did at the start of the procedure, anticipating potential advert events, as well as during the whole procedure. Both novices and experts mobilized energy and proved high level of task engagement. However, as shown in the two cases of surgeons A and B, the same energy mobilization is used for different purpose. When considering the wellbeing of surgeons, it is important to realize that long procedures are definitively more taxing than short ones. Surgeons deploy more mental and physiological resources in such context. In stressful situation the body expends energy resources as an attempt to maintain its equilibrium [9]. Interestingly, conversations that are irrelevant to patients mostly have the role to decrease the stress level. However, irrelevant communications occurring simultaneously with highs in task engagement (GSR) correspond to a delay effect of the surgeon hyper focus. These are the verbal signs of the resolution of the problem (i.e., translating into communicative behaviors). Regarding distractions we surely learned that the accumulation or repetitive annoying sounds are increasing the level. These distractions can really get to the nerves of the surgeons when these obviously add up i.e. the one time opening of a door will not exhaust the surgeon's resources, the repetition and association with a leaking trocar may.

We recognize the limitations of this pilot study. First, only eight surgeons took part to this pilot, and the majority were experts. It will be interesting to involve surgeons from other Dutch hospital in other trials. Second, novice surgeons were not eligible to perform long and complex surgery. We therefore could not compare the level of expertise on the length of the operation. Third, we only reported in detail for 2 full observations. In the future we intend to shadow more operations and propose a better coding of each distractions and interruptions. Indeed, the observations were collected systematically by an observer. In the future it will be interesting to use video-aided observation to increase the reliability of the observation as a form of manipulation check [12]. Finally, instrument changes have been reported as a distraction but is part of procedure flow rather than disruption. However, it can cause disruptions when the wrong instrument is selected [19]. It will be interesting to address the impact of ergonomics factors on the operative flow disruption in detail. Finally, one may argue that some of the external distractions and interruptions are minimum and not as stressful as a major surgical flow disruption. It is indeed a challenge to assess how disturbing a factor is to surgeons. More research is required.

CONCLUSION

This pilot study addressed the effects of surgical flow disruptions on the surgeons' physiological resources from both a quantitative and qualitative perspectives. Disruptions and interruptions of the surgical flow disruption has mostly been addressed from a quantitative perspective. This research underlines the importance to consider the effects of such disruptions on the surgeons' pool of resources. Also, it demonstrates that physiological markers are interesting measurements to assess the disruptive nature of interruption and distraction in the OR. Finally, interest is growing on the potential of virtual immersive training in the medical field [39, 40]. However, little is reported on the importance of realistic team resource management programs in healthcare. Such programs are in widespread use in the military and aviation industries [41, 42]. They are based on simulation and provide training for technical and non-technical skills such as communication and teamwork [43]. As surgeons cannot operate in a bubble, they should not be trained in one [30]. Training is crucial to handle crisis in the OR. As Weenk et al. underlined trainees may benefit recognizing stressors and stressful situations real time and learning to cope with or prevent stress [26]. Training 'in situation', representing more realistically the demands imposed on the surgeons during clinical practice is required to optimize patient safety and preserve surgeons' resources essential to the surgical task. As previously underlined, in "situation" should include disruption of the surgical flow as it repetitively occurs in the OR. Trainees should experience before entering the OR, interruption of their mental flow, competition for their attentional resources, increase level of irritation, while performing surgery on the simulator. Data collected through our observations and then triangulated with physiological markers of body temperature should allow in the future developing realistic scenario, testing in a realistic environment surgeons' nerves of steel.

REFERENCES

[1] R. Berguer, W. Smith, and Y. Chung, *Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery,* Surgical Endoscopy **15**, 1204 (2001).

[2] J. J. Jakimowicz and A. Cuschieri, *Time for evidence-based minimal access surgery training—simulate or sink,* Surgical Endoscopy **19**, 1521 (2005).

[3] B. Zheng, M. A. Cassera, D. V. Martinec, G. O. Spaun, and L. L. Swanström, *Measuring mental workload during the performance of advanced laparoscopic tasks,* Surgical Endoscopy **24**, 45 (2009).

[4] C. G. L. Cao, M. Zhou, D. B. Jones, and S. D. Schwaitzberg, *Can surgeons think and operate with haptics at the same time?* Journal of Gastrointestinal Surgery **11**, 1564 (2007).

[5] C. M. Balch, *Stress and burnout among surgeons,* Archives of Surgery **144**, 371 (2009).

[6] N. Bitterman, *Technologies and solutions for data display in the operating room,* Journal of Clinical Monitoring and Computing **20**, 165 (2006).

[7] B. Zheng, E. Rieder, M. A. Cassera, D. V. Martinec, G. Lee, O. N. M. Panton, A. Park, and L. L. Swanström, *Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures,* Surgical Endoscopy **26**, 1352 (2011).

[8] S. E. Hobfoll and J. Freedy, *Conservation of resources: A general stress theory applied to burnout,* in *Professional Burnout* (Routledge, 2017) pp. 115–129.

[9] Rutkowski, *Emotional and cognitive overload : the dark side of information technology* (Routledge, London, 2019).

[10] D. Kahneman, *Attention and effort* (Prentice-Hall, Englewood Cliffs, N.J, 1973).

[11] L. Monetta and Y. Joanette, *Specificity of the right hemisphere's contribution to verbal communication: The cognitive resources hypothesis,* Journal of Medical Speech-Language Pathology **11**, 203 (2003).

[12] B. Zheng, D. V. Martinec, M. A. Cassera, and L. L. Swanström, *A quantitative study of disruption in the operating room during laparoscopic antireflux surgery,* Surgical Endoscopy **22**, 2171 (2008).

[13] D. A. Wiegmann, A. W. ElBardissi, J. A. Dearani, R. C. Daly, and T. M. Sundt, *Disruptions in surgical flow and their relationship to surgical errors: An exploratory investigation,* Surgery **142**, 658 (2007).

[14] D. Xiao, *Ergonomic Factors during Laparoscopic Surgery Training,* Ph.D. thesis (2014).

[15] G. Palmer, J. H. Abernathy, G. Swinton, D. Allison, J. Greenstein, S. Shappell, K. Juang, and S. T. Reeves, *Realizing improved patient care through human-centered operating room design,* Anesthesiology **119**, 1066 (2013).

[16] T. N. Cohen, J. S. Cabrera, O. D. Sisk, K. L. Welsh, J. H. Abernathy, S. T. Reeves, D. A. Wiegmann, S. A. Shappell, and A. J. Boquet, *Identifying workflow disruptions in the cardiovascular operating room,* Anaesthesia **71**, 948 (2016).

[17] E. Sutton, Y. Youssef, N. Meenaghan, C. Godinez, Y. Xiao, T. Lee, D. Dexter, and A. Park, *Gaze disruptions experienced by the laparoscopic operating surgeon,* Surgical Endoscopy **24**, 1240 (2009).

[18] E. Pilke, *Flow experiences in information technology use,* International Journal of Human-Computer Studies **61**, 347 (2004).

[19] L. Al-Hakim, J. Xiao, and S. Sengupta, *Ergonomics perspective for identifying and reducing internal operative flow disruption for laparoscopic urological surgery,* Surgical Endoscopy **31**, 5043 (2017).

[20] A. M. Tollner, M. A. Riley, G. Matthews, and K. D. Shockley, *Divided attention during adaptation to visual-motor rotation in an endoscopic surgery simulator,* Cognition, Technology & Work **7**, 6 (2005).

[21] C. P. Primus, A. N. Healey, and S. Undre, *Distraction in the urology operating theatre,* BJU International **99**, 493 (2007).

[22] R. Berguer, *A comparison of the physical effort required for laparoscopic and open surgical techniques,* Archives of Surgery **138**, 967 (2003).

[23] J. R. Pluyter, S. N. Buzink, A.-F. Rutkowski, and J. J. Jakimowicz, *Do absorption and realistic distraction influence performance of component task surgical procedure?* Surgical Endoscopy **24**, 902 (2009).

[24] L. Shambo, T. Umadhay, and A. Pedoto, *Music in the operating room: is it a safety hazard?* AANA J **83**, 43 (2015).

[25] J. B. Sexton, *Error, stress, and teamwork in medicine and aviation: cross sectional surveys,* BMJ **320**, 745 (2000).

[26] M. Weenk, A. P. Alken, L. J. Engelen, S. J. Bredie, T. H. van de Belt, and H. van Goor, *Stress measurement in surgeons and residents using a smart patch,* The American Journal of Surgery **216**, 361 (2018).

[27] I. Hassan, P. Weyers, K. Maschuw, B. Dick, B. Gerdes, M. Rothmund, and A. Zielke, *Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance,* British Journal of Surgery **93**, 1554 (2006).

[28] S. Arora, N. Sevdalis, D. Nestel, M. Woloshynowych, A. Darzi, and R. Kneebone, *The impact of stress on surgical performance: A systematic review of the literature,* Surgery **147**, 318 (2010).

[29] C. Powell, B. P. Carson, K. P. Dowd, and A. E. Donnelly, *The accuracy of the SenseWear pro3 and the activPAL3 micro devices for measurement of energy expenditure,* Physiological Measurement **37**, 1715 (2016).

[30] J. R. Pluyter, A.-F. Rutkowski, and J. J. Jakimowicz, *Immersive training: breaking the bubble and measuring the heat,* Surgical Endoscopy **28**, 1545 (2014).

[31] M. E. Dawson, A. M. Schell, D. L. Filion, and G. G. Berntson, *The electrodermal system,* in *Handbook of Psychophysiology*, edited by J. T. Cacioppo, L. G. Tassinary, and G. Berntson (Cambridge University Press) pp. 157–181.

[32] A. Pecchinenda, *The affective significance of skin conductance activity during a difficult problem-solving task,* Cognition & Emotion **10**, 481 (1996).

[33] Y. H. Chan, *Biostatistics 101: data presentation,* Singapore Med J **44**, 280 (2003).

[34] T. P. Hettmansperger and J. McKean, *Robust nonparametric statistical methods,* Statistics in Medicine **18**, 1899 (1999).

[35] S. Levine, *Developmental determinants of sensitivity and resistance to stress,* Psychoneuroendocrinology **30**, 939 (2005).

[36] M. P. Schijven and W. A. Bemelman, *Problems and pitfalls in modern competency-based laparoscopic training,* Surgical Endoscopy **25**, 2159 (2011).

[37] M. P. Schijven, J. J. Jakimowicz, I. A. M. J. Broeders, and L. N. L. Tseng, *The Eindhoven laparoscopic cholecystectomy training course—improving operating*

room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum, Surgical Endoscopy **19**, 1220 (2005).

[38] A. G. Gallagher, G. Leonard, and O. J. Traynor, *Role and feasibility of psychomotor and dexterity testing in selection for surgical training,* ANZ Journal of Surgery **79**, 108 (2009).

[39] M. A. Hagiwara, P. Backlund, H. M. Söderholm, L. Lundberg, M. Lebram, and H. Engström, *Measuring participants' immersion in healthcare simulation: the development of an instrument,* Advances in Simulation **1** (2016), 10.1186/s41077-016-0018-x.

[40] P. Backlund, H. M. Söderholm, H. Engström, M. A. Hagiwara, and M. Lebram, *Breaking out of the bubble putting simulation into context to increase immersion and performance,* Simulation & Gaming **49**, 642 (2018).

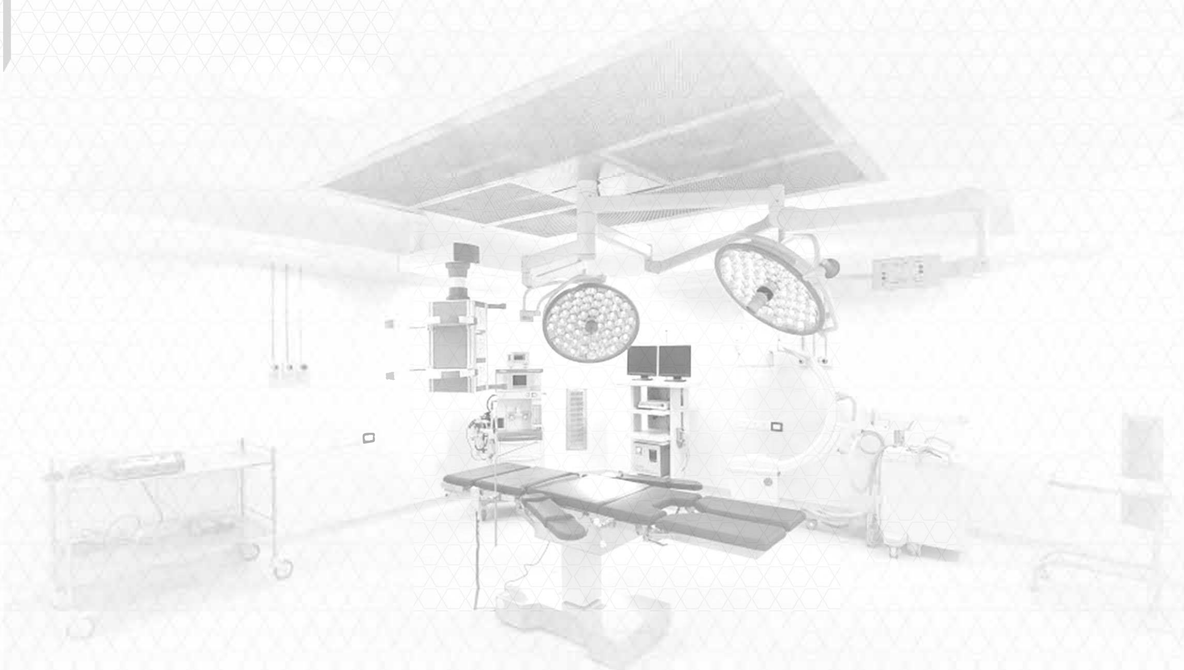
[41] S. Undre, M. Koutantji, N. Sevdalis, S. Gautama, N. Selvapatt, S. Williams, P. Sains, P. McCulloch, A. Darzi, and C. Vincent, *Multidisciplinary crisis simulations: The way forward for training surgical teams,* World Journal of Surgery **31**, 1843 (2007).

[42] S. Yule, R. Flin, S. Paterson-Brown, and N. Maran, *Non-technical skills for surgeons in the operating room: A review of the literature,* Surgery **139**, 140 (2006).

[43] R. Flin, P. O'Connor, and K. Mearns, *Crew resource management: improving team work in high reliability industries,* Team Performance Management: An International Journal **8**, 68 (2002).

10

11



VIRTUAL OPERATING ROOM SIMULATION SETUP (VORSS) FOR PROCEDURAL TRAINING IN MINIMALLY INVASIVE SURGERY – A PILOT STUDY

S.Ganni, M.Li, S.M.B.I.Botden, S.Nayak, B.R.Ganni, A.F.Rutkowski, R.H.M.Goossens, J.J.Jakimowicz

Published in Indian Journal of Surgery, March 2020

11

Abstract

Background

Virtual Reality (VR) training is widely used in several Minimal Invasive Surgery (MIS) training curricula for procedural training. However, VR training in its current state lack immersive training environments, such as using head mounted displays that is implemented in military or aviation training and even entertainment. The Virtual Operating Room Simulation Setup (VORSS) is explored in this study to determine the effectiveness of immersive training in MIS.

Methods

Twenty-eight surgeons and surgical trainees performed a laparoscopic cholecystectomy on the VORSS comprising of a head mounted 360-degree realistic OR surrounding on a VR laparoscopic simulator. The VORSS replicated a full setup of instruments and surgical team-members as well as some of the distractions occurring during surgical procedures. Questionnaires were followed by semi-structured interviews to collect the data.

Results

The participants found the VORSS to be intuitive and easy to use (p = 0.001). The outcome of the usability test, applying QUESI and NASA-TLX, reflected a significant the usability of the VORSS ($p<0.05$), at the cognitive level, which indicates a good sense of

immersion and satisfaction, when performing the procedure within VORSS. The need for personalized experience within the setup was strongly noted from most of the participants.

Conclusions

The VORSS for procedural training has the potential to become a useful tool to provide immersive training in MIS surgery. Further optimizing of the VORSS realism and introduction of distractors in the OR should result in an improvement of the system.

Introduction

Minimally Invasive Surgery (MIS) is rapidly becoming the standard of treatment for many surgical pathologies. [1] However, the skills required to perform MIS are significantly different to that of open surgery. The surgeon has to cope with restricted movement and visual field, fulcrum effect, hand-eye coordination and ever-changing instruments and equipment. [2] Training surgeons to adapt to these challenges requires equally advanced tools that replicate them.

Historically, MIS training has adapted techniques from other fields of technology mostly notably from aviation training. [3] Virtual Reality (VR) simulation has been the cornerstone of training pilots in flight simulation training in that it offers immersive visual and physical representation and replication of real-world scenarios. [4] This has been possible with the use of mock cockpits that are fitted with screens in place of windows and actuators that move the enclosure around making it true to a real-life setting. [5] However, VR simulation in MIS training have not truly achieved the immersion that their counterparts offer.

Current VR simulators for MIS training are equipped with a monitor and instrument handles and foot pedals to perform procedure-specific tasks that replicate tissuespecific haptic feedback. [6] Several validation studies demonstrate the effective transfer of technical skills from the skills labs to the operating room (OR) with the use of procedural VR simulators. [7–9] However, a major deficiency of the current procedural VR simulation is its distraction-void and therefore lack of immersive environments. They are set-up in isolated skills labs or rooms where they seldom replicate the busy and often chaotic operating room (OR) environment. As Pluyter et al state "surgeons cannot operate in a bubble and thus should not be trained in one". [10] It is vital that surgeons are trained in circumstances that replicate the real OR environments. Training in environments that replicate distractions increases the mental load and stress level of the surgeons and helps surgical trainees to adapt faster to the real OR environment. [11]

Distractions that occur during the surgical procedure have been identified and broadly classified into environmental factors, social factors, equipment factors and organizational factors. [12] These range from procedural distractions, such as changing instruments, procedure related conversation between teams, to social factors, such as music, non-procedure related conversations etc. Nowadays available VR headsets have made it accessible and affordable to create immersive environments that replicate true to life with distractions and a sense of being. [13] The combination of VR simulators and VR headsets for the purpose of Virtual Operating Room Simulation Setup (VORSS) for procedural simulation will be explored in this study. We aim to analyze the experience of VORSS by surgeons and surgical trainees and the potential added benefit to the existing procedural VR simulation.

METHODS

PARTICIPANTS

The aim was to include all surgeons and surgical residents from GSL Medical College, Rajahmundry, India to participate in this study. All the participants had prior experience either in real MIS surgery or in using laparoscopic VR simulator or box trainers, laparoscopic instruments and equipment. They were divided into two groups based on their professional background: novices consisted of the surgical residents and the experts were made up of the surgeons. This was based on the demographics questions on the questionnaires completed by the participants.

A total of 28 participants enrolled in the study, of which 15 were residents and 13 were surgeons. Throughout this article we refer to the residents as “novices” and surgeons as “experts”. Of the experts in this study, four had completed >200 cases, three 101-200, three 50-100 and three had performed <50 clinical procedures. Of the novices, 14 had performed fewer than 50 clinical procedures previously and one performed none.

VIRTUAL OPERATING ROOM SIMULATION SETUP (VORSS)

The VORSS contains three essential components: a VR laparoscopic simulator (1), a VR headset (2) and a virtual OR environment (3).

The VR laparoscopic simulator (1): LapMentor III (Simbionix™, 3D Systems Corporation, the US) with MentorLearn Software. The specific hardware includes a 24" flat touch-screen monitor, a keyboard with trackball, two instrument handles offering tactile feedback, and a double footswitch for activating simulated electrosurgical coagulation.

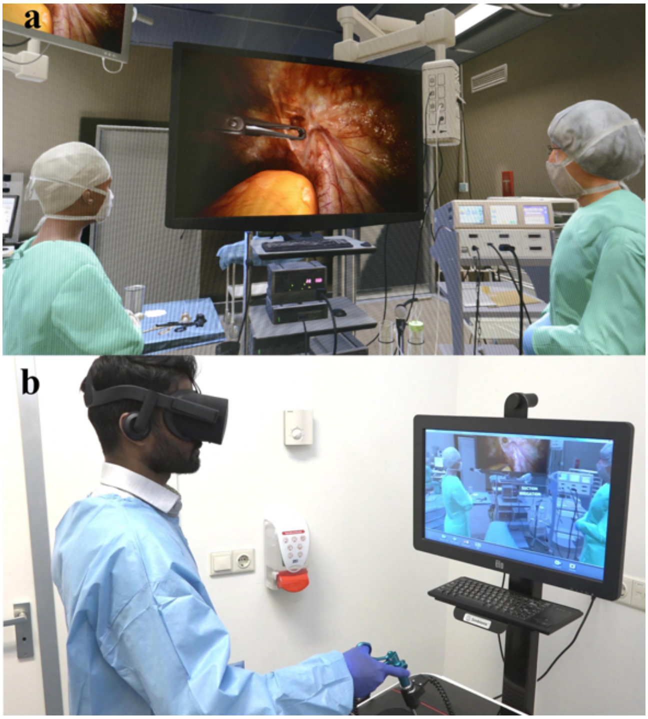


Figure 11.1: a: the replicated OR setup of the VORSS. b: an external view of the setup of the VORSS.

The VR headset (2): 2016 Oculus Rift providing stereoscopic images (1080*1200 per eye, 110°field of view), integrated 3D audio and six degree-of-freedom head-tracking.

The virtual OR environment (3): a panoramic VR scene regenerates a real OR including a full setup of instruments and equipment, and as a new feature, also a surgical team and various distractions. The distractions cover some of the distractive events observed in a real OR [14] (Figure 11.1a). The virtual OR can be simultaneously seen on the monitor and in the VR headset from the same point of view (Figure 11.1b).

TASK

Firstly, the purpose was introduced to the participants to the VORSS system, to evaluate the use of VORSS in procedural VR simulation training in a realistic OR context. Participants were introduced to the VORSS and given time to familiarize themselves with the system. Informed consent was completed by the participants before the start of the study.

After the participants put on the VR headset, the VR simulator was adjusted ergonomically according to their height. Then they started a hands-on task “Complete Laparoscopic Cholecystectomy Procedure”, which was previously validated as a basic procedural module of Laparoscopic Surgical Skill Grade 1 Level 1 course [15]. A predefined protocol required participants to interact with the VORSS for 15 minutes. Since the task not aimed at assessing their performance, participants could stop whenever they thought it was enough to evaluate the VORSS.

After completing the task, the participants were asked to complete four questionnaires related to the VORSS experience. At the end, general suggestions and comments could be made regarding the realism of the VORSS by participants.

ASSESSMENT METHODS

The participants were asked to score questions regarding the immersion, usability and reality of the VORSS experience. Since this is an efficacy study, power calculations were not performed a priori. While our sample size is small, one of the strengths of our approach in this study is that we present the results of multiple validated tools to assess each criterion. [16] The responses were analyzed via Presence Questionnaire (PQ) [17], Questionnaire for Intuitive Use (QUESI) [18], NASA-Task Load Index (NASA-TLX) [19], and a heuristics questionnaire.

The Presence Questionnaire was modified and previously validated (Cronbach a = 0.878) to measure the immersion at sensory level [20, 21]. The PQ contained twenty-four items reflecting seven influencing factors for self-reported immersion, including

Realism, Possibility to act, Quality of interface, Possibility to examine, Self-evaluation of performance, along with haptic and sound factors. The study added two items on haptic and one item on sound according to the VORSS. An extended 7-point scale was used in fine gradient in which 1 is not immersive and 21 completely [22]. A baseline of the high level of immersion was assigned as 15 [18].

The Questionnaire for Intuitive Use (QUESI) indicated the subjective satisfaction of interacting with the immersive VORSS [18]. The QUESI measures five aspects of satisfaction using a 5-point Likert scale. The baselines of the subscales and total were set respectively according to Hurtienne and Naumann [17].

The NASA-TLX assessed the mental workload or performance problem when performing the task in VORSS [19, 23]. The sub-scales measured six factors of the mental effort from very low (1) to very high (21). A baseline value was assigned as 11 represented a medium level of workload.

A questionnaire was developed based on the ease-of-use heuristics for medical devices. [24] Participants used the heuristics as a guideline to rate their experience with a 5-point scale at system level, in which 1 means not realistic and 5 completely. A baseline of reality was considered as 4 , indicating that only appearance problems were encountered by participants when using the VORSS.

As the final step of the assessment, participants were interviewed with two questions: (1) How satisfied are you with the Virtual OR experience? (2) Which factors were not compelling or not realistic in the Virtual OR experience?

Statistical Analysis

The data was analyzed using SPSS v.25. The mean and standard deviation of each questionnaire of the sample, novices and experts were calculated as well as the median and interquartile range (IQR) . The means and the baselines were then compared using one-sample t-test (normally distributed) or Wilcoxon Signed Rank test (non-normally distributed). The differences between novices and experts were tested using classical Independent-sample t-test, otherwise non-parametric tests such as the Kruskal-Wallis test and Mann-Whitney U test where appropriate.

Results

Participants

A total of 28 participants enrolled in the study, of which 15 were novices (surgical residents) and 13 were experts (surgeons). Of the experts in this study, four had completed >200 cases, three 101-200, three 50-100 and three had performed <50 clinical procedures. Of the novices, 14 had performed fewer than 50 clinical procedures previously and one performed none. There were 8 male 7 female novices and 9 male and 4 female experts.

Immersion: Presence Questionnaire

Figure 11.5 presents the results of the self-reported immersion from the subscales of the Presence Questionnaire. In summary, the four subscales - Realism, Possibility to act, Quality of interface and Haptic - as well as the overall total had a significantly lower level of immersion than the baseline (PQ subscales= 15, $p < .05$). Both novices and experts had similar immersion level across the subscales and overall, which were also all significantly different from the threshold. There were no significant differences between the opinion of the novices and experts.

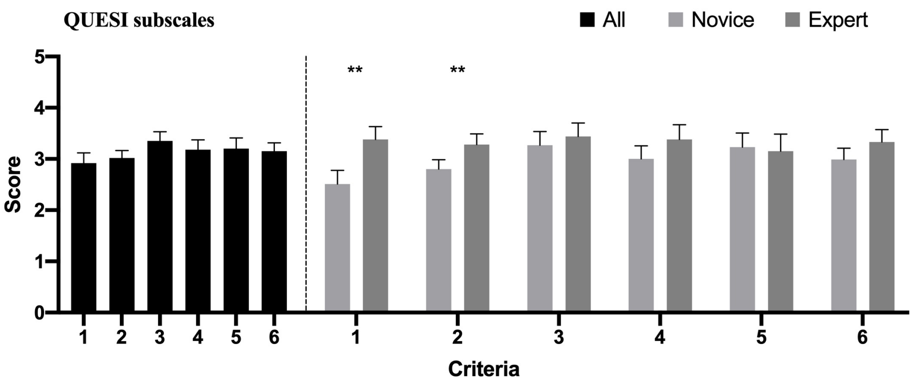
Usability: QUESI and NASA-TLX

The QUESI and NASA-TLX both reflected the usability of VORSS at cognitive level. The five subscales and total score of the QUESI were calculated to discover whether the participants were satisfied when performing the task within the VORSS (Figure 11.2). None of the subscales nor the total score of VORSS were significantly lower than the baselines ($W = 2.94$, $G = 2.89$, $L = 3.00$, $F = 2.88$, $E = 3.04$, total = 2.95). However, the subjective mental workload and perceived achievement of goals for VORSS were significantly lower for the novices than the experts ($p < 0.05$).

Six subscales of NASA-TLX were calculated to detect the main sources of mental workload (Figure 11.3). The mental demand was significantly higher than the baseline, while frustration and performance were significantly lower than it (NASA-TLX subscales = 11, $p < 0.05$). In addition, the novices had a significantly higher mental workload in mental demand than experts ($p < 0.05$).

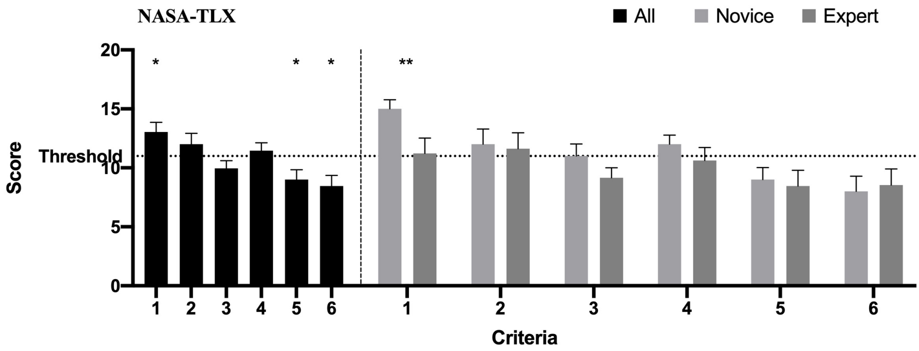
Reality: Heuristics questionnaire

Fourteen heuristics were analysed to judge the reality of VORSS at system level. Figure 11.4 shows the criteria of the heuristics instead of the full guideline. All fourteen heuristics scored significantly lower than the baselines (heuristics=4, p<0.05). The experts showed significantly higher agreement on the heuristic the VORSS Prevent errors and Reversible actions categories ($p < 0.05$) than the novices did.



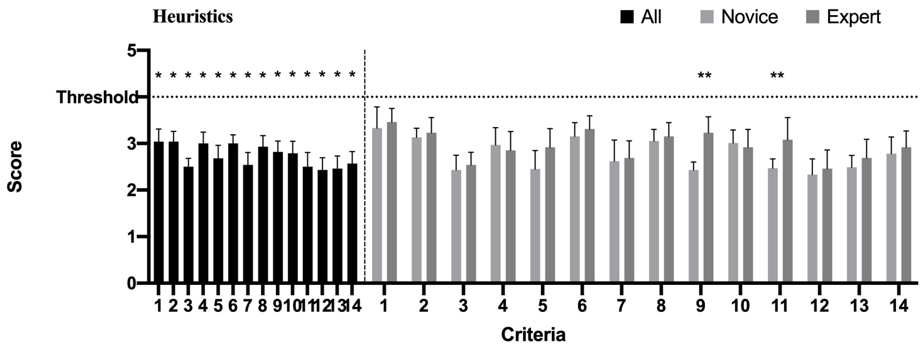
QUESI subscales (items)		1 Subjective mental workload (W) (1,6,11)	2 Perceived achievement of goals (G) (2,7,12)	3 Perceived effort of learning (L) (3,8,13)	4 Familiarity (F) (4,9,14)	5 Perceived error rate (E) (5,10)	6 Total
Novice	Mean (SD)	**2.51** (1.03)**	**2.80 ** (0.72)**	3.27 (1.03)	3.00 (1.00)	3.23 (1.07)	2.99 (0.86)
Expert	Mean (SD)	**3.38** (0.9)**	**3.28** (0.76)**	3.44 (0.95)	3.38 (1.04)	3.15 (1.21)	3.33 (0.87)
ALL	Mean (SD)	2.92 (-1.05)	3.02 (0.76)	3.35 (0.97)	3.18 (1.02)	3.2 (1.12)	3.15 (0.86)
	P_mean (2-tail)	0.91	0.36	0.07	0.13	0.46	0.24

Figure 11.2: Table and graph showing summary data for the level of intuitive use of the VORSS with (**) indicating significant ($p < 0.05$) difference between the mean for novices and experts. (1 = "Fully disagree", 5 = "Fully agree"))



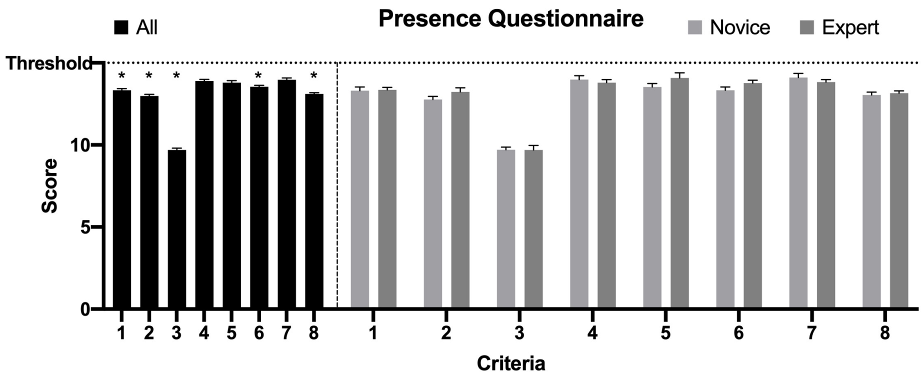
NASA-TLX		1 Mental Demand	2 Physical Demand	3 Temporal Demand	4 Effort	5 Frustration	6 Performance
Novice	Mean (SD)	**15.01** (3.03)**	12.33 (5.05)	11.15 (4.15)	12.16 (3.02)	9.47 (3.85)	7.96 (4.87)
Expert	Mean (SD)	**11.23** (4.66)**	11.62 (4.91)	9.15 (3.11)	10.62 (3.99)	8.46 (4.81)	8.54 (4.94)
ALL	Mean (SD)	13.04 (4.32)	12 (4.90)	9.96 (3.47)	11.46 (3.52)	9 (4.42)	8.46 (4.77)
	P_mean (2-tail)	**0.02***	0.29	0.13	0.49	**0.02***	**0.01***

Figure 11.3: Table and graph showing summary data for the self-reported mental workload after using the VORSS with (*) indicating significant ($p < 0.05$) difference between the mean for the whole data set and the threshold and (**) indicating significant ($p < 0.05$) difference between the mean for novices and experts. (1 = "Very low", 21 = "Very high"))



Heuristics	Novice	Expert	All	
	Mean (SD)	Mean (SD)	Mean (SSD)	P mean (2 tail)
1 Consistent and Standardized	3.33 (1.76)	3.46 (1.05)	3.04 (1.43)	**<0.001**
2 Visible	3.13 (0.76)	3.23 (1.17)	3.04 (1.17)	**<0.001**
3 Match with real world	2.43 (1.23)	2.54 (0.97)	2.5 (0.96)	**<0.001**
4 Minimalist	2.97 (1.43)	2.85 (1.46)	3 (1.28)	**<0.001**
5 Minimizes Memory Load	2.45 (1.54)	2.92 (1.44)	2.68 (1.49)	**<0.001**
6 Informative Feedback	3.15 (1.15)	3.31 (1.03)	3 (0.98)	**<0.001**
7 Flexible and Efficient	2.62 (1.76)	2.69 (1.32)	2.54 (1.4)	**<0.001**
8 Good Error messages	3.05 (0.98)	3.15 (1.07)	2.93 (1.27)	**<0.001**
9 Prevent Errors	**2.43 (1.07)**	**3.23* (1.24)**	2.82 (1.22)	**<0.001**
10 Clear Closure	3.01 (1.09)	2.92 (1.38)	2.79 (1.37)	**<0.001**
11 Reversible Actions	**2.47 (0.86)**	**3.08* (1.71)**	2.5 (1.62)	**<0.001**
12 Uses users' language	2.33 (1.32)	2.46 (1.45)	2.43 (1.4)	**<0.001**
13 Users control	2.49 (0.98)	2.69 (1.44)	2.46 (1.43)	**<0.001**
14 Help and Documentation	2.78 (1.39)	2.92 (1.26)	2.57 (1.37)	**<0.001**

Figure 11.4: Table and Graph showing summary data for the level of reality of the VORSS with (*) indicating significant (p<0.05) difference between the mean for the whole data set and the threshold and (**) indicating significant (p<0.05) difference between the mean for novices and experts. (1= fully disagree, 5=fully agree))



Presence Questionnaire		1 Realism	2 Possibility to Act	3 Quality of interface	4 PTE	5 SEOP	6 Haptic	7 Sound	8 Total
Novice	Mean (SD)	13.30 (0.89)	12.77 (0.72)	9.71 (0.6)	13.98 (0.91)	13.53 (0.82)	13.33 (0.79)	14.10 (0.99)	13.04 (0.68)
Expert	Mean (SD)	13.35 (0.56)	13.23 (0.91)	9.69 (0.99)	13.79 (0.67)	14.08 (1.13)	13.77 (0.61)	13.83 (0.53)	13.16 (0.5)
All	Mean (SD)	13.33 (0.53)	12.98 (0.56)	9.7 (0.55)	13.89 (0.57)	13.79 (0.67)	13.54 (0.5)	13.97 (0.58)	13.1 (0.43)
	P-value (2-tailed)	**.004***	**.001***	**<.001***	0.061	0.082	**.007***	0.085	**<.001***

Figure 11.5: Table and graph showing summary data for self-reported immersion from the subscales of Presence Questionnaire with (*) indicating significant (p<0.05) difference between the mean for the whole data set and the threshold (1="Not at all", 11="Somewhat", 21="Completely")

SEMI-STRUCTURED INTERVIEW

Comments solicited from the participants were broadly categorized into Virtual OR experience related, OR team related and Personalization related:

VIRTUAL OR EXPERIENCE

Participants were critical on a few aspects of the VOR experience pertaining interaction between the VOR and the VR simulator. Some could not see their own legs and foot pedals because the system did not allow them. Some comments were related to the procedural steps of the laparoscopic cholecystectomy depicted in the VR simulator perceived being different from their way of practice. Overall the participants were intrigued with the novelty of the system and were proactive in using and validating the system.

OR team

Several participants commented on OR team and how it affected their perception of level of system realism. The team would normally be located different to the placement depicted in the VORSS. The team spoke English as opposed to the local language. The interaction between the team is not realistic and distracting. The voices in the background were unfamiliar and unrelated. The aggregate perception towards the OR team reproduction was negative

Personalization

Overall the participants felt the system could benefit from personalization to meet individual preferences and realistic workplace replication.

Discussion

VR simulators have been successfully implemented to different training curricula in MIS, significantly contributing to acquisition of skills, which is mandatory for a safe performance of MIS surgery. [25] The outcome of multiple validation studies of VR simulators indicate that they adequately reproduce the surgical procedures, operative techniques and instrumentation. [26] This has proven to be of value in providing constant objective evaluation of the task and procedural performance. The challenges of current VR simulators and simulation settings face lack of the system realism and immersion that are otherwise present in other fields of simulation training, such as in aviation, military training and even in the entertainment.

The VORSS outlined and validated in this study builds upon the strength of the VR procedural simulation, and provides additional immersion experience of the operating room. The outcome of the usability, by applying QUESI and NASA-TLX tests, reflect the usability of the VORSS, at the cognitive level, which indicates a good sense of immersion and satisfaction, when performing the procedure within VORSS. The difference in mental workload was perceived significantly different by experts than novices, indicating that performing the task itself was more demanding for the surgical residents (novices) that the more experienced surgeons (experts). Increased mental load created by the VOR environment with additional distractions and tasks, with the introduction of the OR team, implicates that trainees will be better prepared and will adapt to the work environment in the real OR more easily and faster. This has been proven in prior research, when exploring the role of the distractors and increased mental load in course of procedural VR training in skills lab setting. [11] The outcome of this study has demonstrated clearly that training in an environment mimicking the real work place shows higher efficiency of training shortening of adaptation period to the real OR environment. Benefits of this approach is demonstrated and proven by using immersive training programs for military personnel, emergency crew training and ICU personnel showing shorter learning curves and shortened adaption period to real world setting. [27, 28]

Regarding the issue of self-assessment from our prior studies we found that self-assessment has a good correlation with expert assessment and VR simulator assessment. [29] However, it is interesting to note that both experts and novices over-assess their performance in this study. While it seems to be possible to over-assess their performance in new immersive training environment [11], it is crucial to develop objective criteria, next to the existing VR simulation criteria, for accurate self-assessment in VORSS setting. Implementing of the self-assessment component within the VORSS could importantly contribute to self-development and proficiency awareness of trainees.

The semi-structured interviews of the participants show a strong emphasis of the user perception on personalization. All users appreciated the immersive environment, created by the VORSS. The lack of personalization pertaining to language, crew placement, crew interaction, instrument-specific personalization, OR-layout were considered to be less realistic. This obviously indicates the need to improve realism of the virtual environment, focusing upon above mentioned aspects. One should also consider potentially customizing the environment, taking into account specific conditions, related to the region of the world, country or even specific institution were training takes place. This approach could lead to optimizing the procedural VR simulation training, resulting in improvement of safety and quality of MIS surgery. Furthermore, with the increased training demands of trainees and trainer constraints in India, there is an imminent need to address these challenges with effective tools that prepare a trainee for the operating room. [30] Future extensions of this work could include a study into the cost-effectiveness of this approach compared with mentor-mentee training, the use of simulated OR experience in a skills lab setting and a multi-national validation study to confirm the effects seen here.

CONCLUSION

The VORSS for procedural training has the potential to become a useful tool to provide immersive training in MIS surgery. Further optimizing of the VORSS improving realism and introduction of distractors in the VOR should result in improvement in the effectiveness of the procedural training by shortening the learning curve and speeding up the adaption of trainees to the real OR setting.

REFERENCES

[1] A. van Dijk, M. Lamberts, C. van Laarhoven, J. Drenth, M. Boermeester, and P. de Reuver, *Laparoscopy in cholecysto-choledocholithiasis,* Best Practice & Research Clinical Gastroenterology **28**, 195 (2014).

[2] R. Eyal and F. Tendick, *Spatial ability and learning the use of an angled laparoscope in a virtual environment,* Stud Health Technol Inform **81**, 146 (2001).

[3] J. B. Cooper and V. R. Taqueti, *A brief history of the development of mannequin simulators for clinical education and training,* Postgraduate Medical Journal **84**, 563 (2008).

[4] I. W. Strachan, *Technology leaps all around propel advances in simulators,* Accessed June 19, 2019.

[5] J. Barton, B. Cybyk, D. Drewry, T. Frey, B. Funk, R. Hawthorne, and J. Keane, *Use of a high-fidelity uas simulation for design, testing, training, and mission planning for operation in complex environments,* (2011).

[6] A. Zhang, M. Hünerbein, Y. Dai, P. M. Schlag, and S. Beller, *Construct validity testing of a laparoscopic surgery simulator (lap mentor®),* Surgical Endoscopy **22**, 1440 (2007).

[7] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, and R. M. Satava, *Virtual reality training improves operating room performance,* Annals of Surgery **236**, 458 (2002).

[8] M. P. Schijven, J. J. Jakimowicz, I. A. M. J. Broeders, and L. N. L. Tseng, *The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum,* Surgical Endoscopy **19**, 1220 (2005).

[9] S. Bann, A. Darzi, Y. Munz, B. D. Kumar, and K. Moorthy, *Laparoscopic virtual reality and box trainers: is one superior to the other?* Surgical Endoscopy **18**, 485 (2004).

[10] J. R. Pluyter, A.-F. Rutkowski, and J. J. Jakimowicz, *Immersive training: breaking the bubble and measuring the heat,* Surgical Endoscopy **28**, 1545 (2014).

[11] J. R. Pluyter, S. N. Buzink, A.-F. Rutkowski, and J. J. Jakimowicz, *Do absorption and realistic distraction influence performance of component task surgical procedure?* Surgical Endoscopy **24**, 902 (2009).

[12] M. C. Persoon, H. J. H. P. Broos, J. A. Witjes, A. J. M. Hendrikx, and A. J. J. M. Scherpbier, *The effect of distractions in the operating room during endourological procedures,* Surgical Endoscopy **25**, 437 (2010).

[13] A. Almeida, F. Rebelo, and P. Noriega, *Development of a virtual environment for safety warnings behavior compliance evaluation,* in *Advances in Ergonomics in Design* (Springer International Publishing, 2016) pp. 35–42.

[14] H. M. Mentis, A. Chellali, K. Manser, C. G. L. Cao, and S. D. Schwaitzberg, *A systematic review of the effect of distraction on surgeon performance: directions for operating room policy and surgical training,* Surgical Endoscopy **30**, 1713 (2015).

[15] S. Buzink, M. Soltes, J. Radonak, A. Fingerhut, G. Hanna, and J. Jakimowicz, *Laparoscopic surgical skills programme: preliminary evaluation of grade i level 1 courses by trainees,* Videosurgery and Other Miniinvasive Techniques **3**, 188 (2012).

[16] Y. Zhang, R. Hedo, A. Rivera, R. Rull, S. Richardson, and X. M. Tu, *Post hoc power analysis: is it an informative and meaningful analysis?* General Psychiatry **32**, e100069 (2019).

[17] A. Naumann and J. Hurtienne, *Benchmarks for intuitive interaction with mobile devices,* in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI '10* (ACM Press, 2010).

[18] S. Deb, D. W. Carruth, R. Sween, L. Strawderman, and T. M. Garrison, *Efficacy of virtual reality in pedestrian safety research,* Applied Ergonomics **65**, 449 (2017).

[19] S. G. Hart, *Nasa-task load index (NASA-TLX); 20 years later,* Proceedings of the Human Factors and Ergonomics Society Annual Meeting **50**, 904 (2006).

[20] B. G. Witmer and M. J. Singer, *Measuring presence in virtual environments: A presence questionnaire,* Presence: Teleoperators and Virtual Environments **7**, 225 (1998).

[21] B. G. Witmer, C. J. Jerome, and M. J. Singer, *The factor structure of the presence questionnaire,* Presence: Teleoperators and Virtual Environments **14**, 298 (2005).

[22] R. F. Dyer, J. J. Mattherws, C. E. Wright, and K. L. Yudowitch, *Questionnaire construction manual,* Accessed 05 April, 2019.

[23] M. Li, A. Albayrak, Y. Zhang, and D. van Eijk, *Multiple factors mental load evaluation on smartphone user interface,* in *Advances in Intelligent Systems and Computing* (Springer International Publishing, 2018) pp. 302–315.

[24] J. Zhang, T. R. Johnson, V. L. Patel, D. L. Paige, and T. Kubose, *Using usability heuristics to evaluate patient safety of medical devices,* Journal of Biomedical Informatics **36**, 23 (2003).

[25] J. J. Jakimowicz and S. Buzink, *Training curriculum in minimal access surgery,* in *Training in Minimal Access Surgery* (Springer London, 2015) pp. 15–34.

[26] S. R. Dawe, G. N. Pena, J. A. Windsor, J. A. J. L. Broeders, P. C. Cregan, P. J. Hewett, and G. J. Maddern, *Systematic review of skills transfer after surgical simulation-based training,* British Journal of Surgery **101**, 1063 (2014).

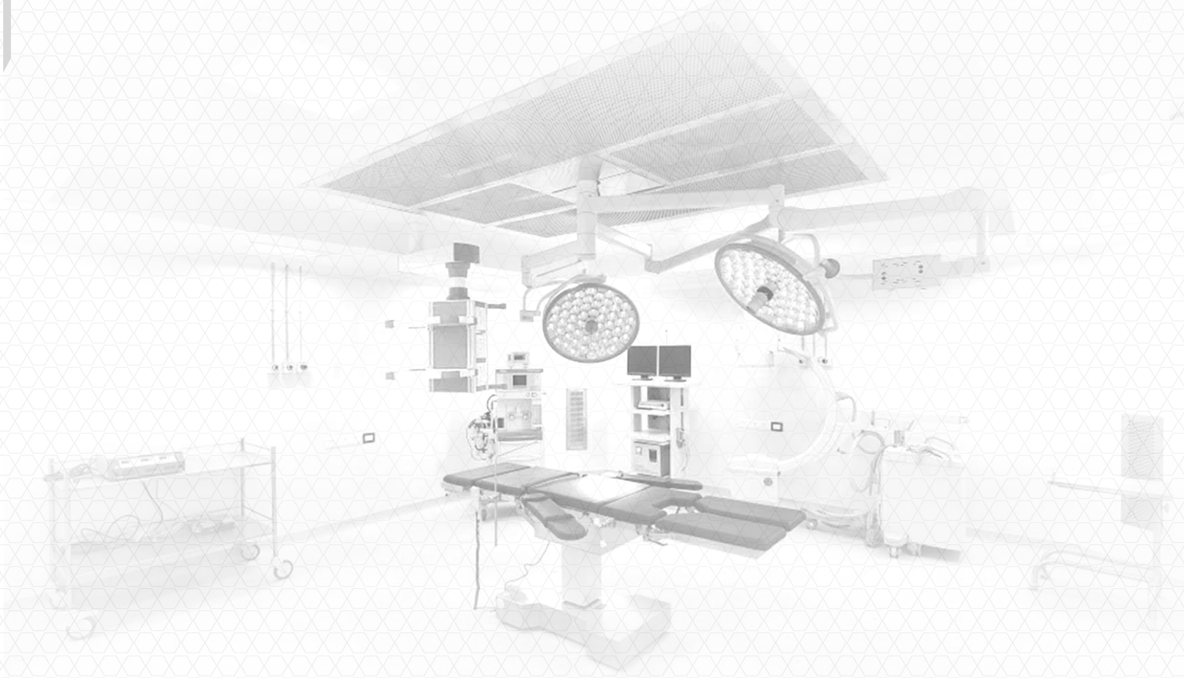
[27] K. K. Bhagat, W.-K. Liou, and C.-Y. Chang, *A cost-effective interactive 3d virtual reality system applied to military live firing training,* Virtual Reality **20**, 127 (2016).

[28] M. F. Hamilton, *Virtual reality,* Pediatric Critical Care Medicine **20**, 587 (2019).

[29] S. Ganni, M. K. Chmarra, R. H. M. Goossens, and J. J. Jakimowicz, *Self-assessment in laparoscopic surgical skills training: Is it reliable?* Surgical Endoscopy **31**, 2451 (2016).

[30] Chintamani and S. Rastogi, *Tool or the teacher?—mentoring and mastering using simulation in surgery,* Indian Journal of Surgery **81**, 207 (2019).

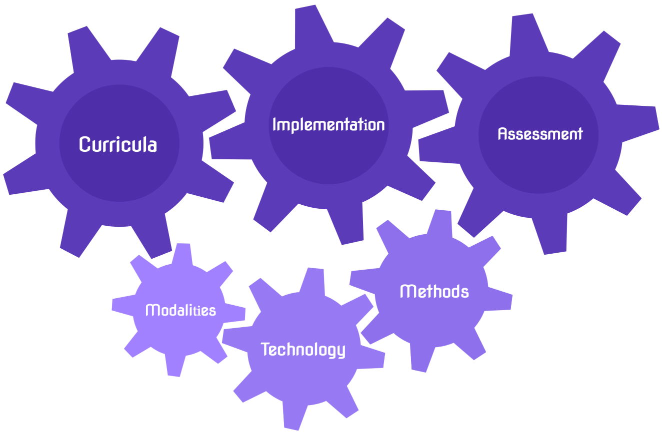
12



DISCUSSION AND FUTURE PROJECTS

12

The premise of this thesis converges upon the core aspects of training in MAS. These can be subdivided into a concentric and interdependent flow as represented below:



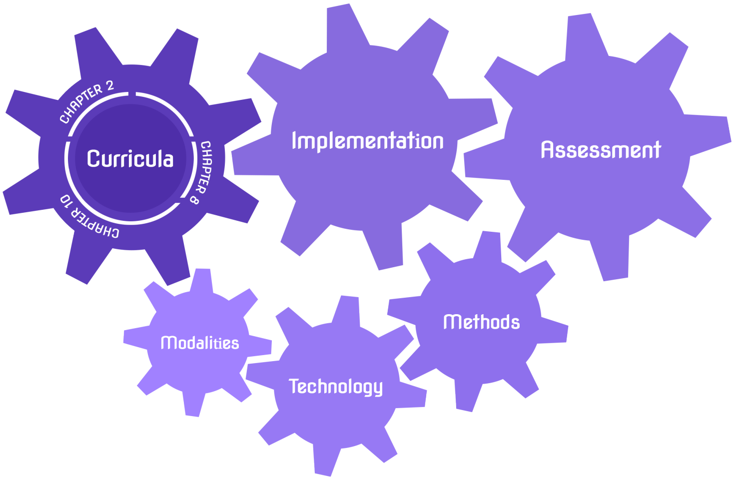
One of the biggest paradigm shifts the field of surgery has seen in the past decades has been the adoption of laparoscopy/minimal access surgery (MAS), developments that are technology driven and independent. [1–3] The advantages of MAS have been extensively documented to affect all the stakeholders involved. [4–7] Ever since, there has been a gradual yet consistent replacement of open procedures to MAS. Several of such procedures performed by MAS are now considered a gold standard of treatment. [8–11] However, as with introduction of any such radically different and technologically dependent procedures, there are bound to be significant challenges in the manner of adoption and training. [12–14]

The successful implementation of training in MAS requires not only the use of validated training tools and curricula but also an organizational and personnel culture that embraces and adapts to change. [15, 16] The aim of this thesis was to identify the key elements that are essential and consequential to adapting to newer training methods. Methods that improve and safeguard the quality and safety of MAS. In this chapter we address these key elements under the aegis of Curricula, Implementation, Assessment, Modalities, Technologies and Methods. The interrelation and progression between these elements as seen in the illustration above can be inferred as crucial for implementation of a training program in MAS.

The first phase of design considered crucial is the Curricula that provides a framework for bringing together the desired elements of training. The second and most important phase is the Implementation phase which makes the curricula effective and brings out the envisioned benefits of that program. The third phase, Assessment sheds an important light on the effectiveness of both the curricula and implementation and provides feedback for improvement of both aspects of the training. The last and fourth phase is the Modalities, Technologies and Methods phase, which aids in improving the design of curricula by using novel tools and the accurate assessment of the participants using such advanced tools and software.

CURRICULA

Throughout the brief history of training in MAS, there has been a considerable misconception of understanding the concept of curriculum. As summarized in Chapter 2, the survey strongly indicated that most of the surgical trainees and surgeons had an incomplete understanding as to what a curriculum encompasses. Often, their understanding was an overlap of different forms of training, such as training courses and workshops onto a curriculum. A curriculum, however, is a culmination of content, educational strategies, educational environment, learning outcomes, learning opportunities and assessment. Whilst major branches of medical education including general surgery are in essence taught under the principles of curriculum design, MAS as



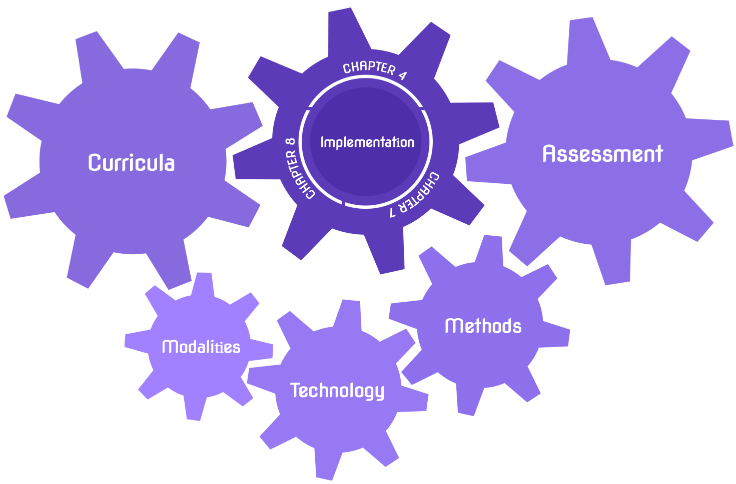
such is not. MAS with its multidisciplinary learning needs and tangential deviation from that of general surgery warrants a curriculum of its own. An important element in curriculum design is the consistent use of validated tools once the learning goals and learning environments are identified. In Chapter 2, the results show, even those who had access to the training courses and certificate programs had varied implementation of training. The tools used in those programs varied significantly from one another and often did not have assessment as part of the certification.

As outlined in Chapter 8, the control group in the three-year study was trained in MAS under the curriculum guidelines of general surgery, which had relatively low emphasis and delineation of MAS from that of general surgery. In comparison, a study group was subject to the LSS curriculum with the sole emphasis of training MAS skills. This in essence led to a greater focus by the participants of the study group on the elements of training in MAS and thus led to significantly better performance outcomes in terms of surgical performance and patient outcomes.

An important point to note when designing a curriculum is the consideration of the hidden curriculum. [17] Regardless of how a curriculum is designed, which is considered to be the declared curriculum, it is taught in practice relative to the methods and practices used by the educators and is considered as the taught curriculum. The hidden curriculum, however, is the actual curriculum learned by the learner. As such, it is broadly subject to individual learning abilities and learning environment indirectly. No matter how well designed a curriculum is, its effectiveness is severely limited if the aspects of hidden curriculum are not taken into account. The consistent redesigning and improving the curriculum upon the feedback and observation of participants results in better learning outcomes as envisioned in the declared curriculum.

Educational environment is considered as one of the core principles to take into consideration when designing a curriculum. [18, 19] Therefore, it is essential to take into consideration the environment where these skills are practiced in. In MAS training it is essential to design curricula taking into consideration the OR environment where several factors such as disruptions to work flow, team coherence and unexpected complications directly affect the surgical performance. In Chapter 10, we address this specific issue by objectifying the disruptions that garner the attention of the surgeons. The use of physiological markers used in the study have provided an insight into what has to be taken into consideration whilst recreating these disruptions and transferring them over to the training environments.

IMPLEMENTATION



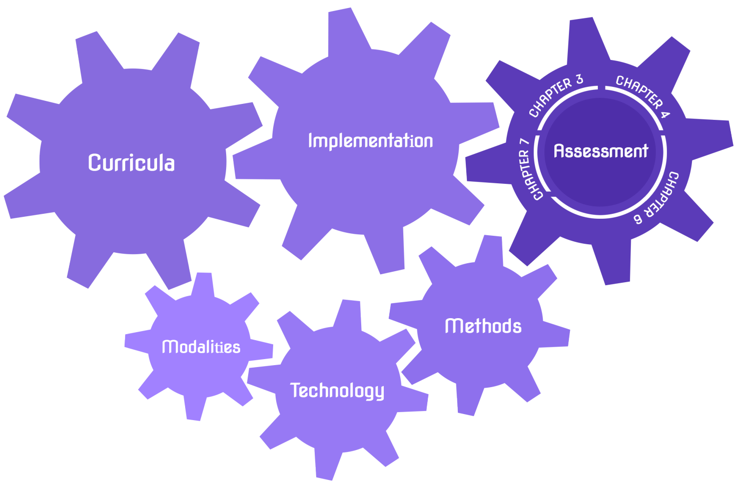
The importance of implementing a curriculum for MAS skills cannot be understated. As discussed in the curriculum section above, there are several stakeholders and resources that need to come together to make up for a successful MAS curriculum. However, having these said resources alone does not guarantee success. There are several instances that we observed across the skills labs in different parts of the world where organizational and process failure led to the resources end up being underutilized and thereby resulting in partial implementation of intended training programs and curricula.

From the experience and feedback gained throughout this research, it became evident that all of the surgical residents we surveyed and those who participated in the study were actively looking for training courses and certification programs to train their MAS skills in their post-residency period. Their main reasons for looking for these courses was their lack of adequate exposure to MAS training during their residency and their lack of confidence in performing MAS. In most cases, they had to assist a senior surgeon during or after their residency and for a number of procedures and optionally attending a MAS training course before performing on their own. In Chapter 8, during the period of the study, the LSS curriculum was integrated into the three-year residency curriculum after identifying the needs of the residents. After six months into their residency program they were taught the basics of MAS as part of the LSS curriculum with access to skills lab training on box trainers and simulators for over a period of eight weeks with mandatory teaching and discussion sessions. After passing the two-day certified program in the skills lab they were given access to perform in the OR under the guidance of a senior surgeon. This training extended well into their third year of residency and their performance was monitored.

The essential task of structural change to the existing general surgery curriculum by means of unequivocal support of the management and the faculty in allocating resources and time required to organize the LSS curriculum was deemed key for successful implementation. Furthermore, the implementation of mandatory requirement of five index laparoscopic procedures to be completed by the residents before graduation, made their active participation imperative. Another interesting element to the implementation was the prospect of certification of their laparoscopic skills by the LSS foundation at the end of their successful completion of the curriculum. This essentially put them at an advantage over their colleagues in other medical schools, which increased their prospects post-graduation as evidenced in Chapter 8. Conversely, for the control group in the study, it was necessary to make arrangements for them to meet the minimum requirements of the laparoscopic procedures due to lack of organizational support from their originating institution. Interestingly, when enquired about this, the reason for noncompliance stemmed from MAS training not being part of the mandated curriculum or as deemed essential by the management.

Similarly, in Chapter 4, taking into consideration the feedback of participants and trainers, and after due observation of the process of self-assessment, we set out to amend how the process of self-assessment is implemented. This required a fundamental review of the concept of self-assessment, deciphering the principles of reflective practice and applying them to assessment practices in MAS training. The changes brought about in the implementation of self-assessment training and changes to evaluation tools used in the curriculum resulted in improved correlation with expert assessment and thus increased overall satisfaction with the curriculum. Furthermore, our study found that using reflective training approach to convey goals of learning and evaluation served as a precursor to improved performance and progression of skills. This indicated the compounding effect of enhancing one aspect of the curriculum had on another and the importance of continuous structural change based on current trends and end-user evaluations.

ASSESSMENT

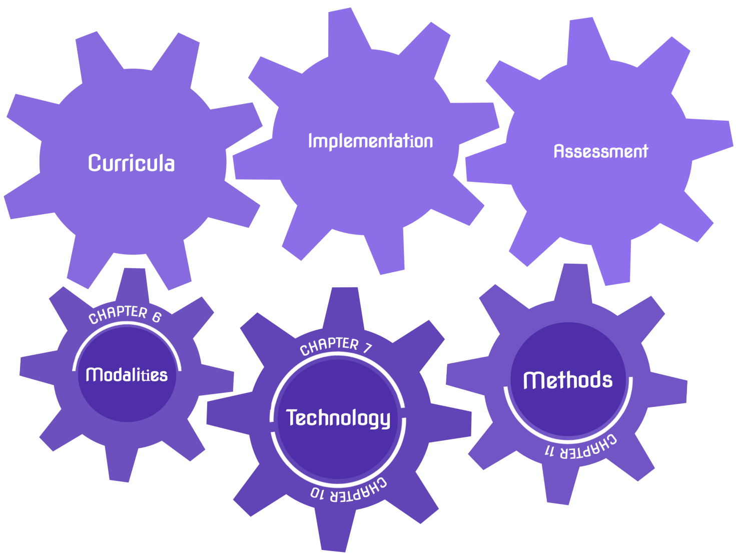


Assessment of skills in MAS training is intricate and labyrinthine and requires the utmost importance when the aspects of using and implementing assessment tools are taken into consideration. As it stands, there are several modes of assessment that serve as indicators of not just proficiency of the trainees but also serve as the indicators of the quality of the mode of training imparted. Moreover, the feedback gained from the effective use of these assessments is self-regulatory for the improvement of practices, progression of skills and aids in a perspective altering insights for assessors, trainees and peers. At the same time, several studies demonstrate the lack of structured assessment leads to over-assessment of skills obtained and even build upon the erroneous techniques learned in the courses in the longer run potentially hindering betterment of surgical skills and patient safety. [20]

In this thesis, the initial global survey conducted on the practices of training in MAS found a fragmented approach to implementing assessment protocols and in some instances even circumvented using assessments. Several, respondents those that comprised of surgeons and surgeons in training recalled scenarios of training courses without assessments or in some cases even had no knowledge of standardized assessment tools. In Chapter 3, the aim of the study was to analyze the role of self-assessment in MAS training. Herein the different multi-modal elements of the LSS curriculum were taken into consideration that included the box-trainers, VR simulators and the tasks simulated within. The tie-in of self-assessment with expert assessment and the correlation between them should be taken into consideration as it is directly proportional to the learning outcomes as envisioned by the curriculum. Scoring better in assessment though one crucial aspect of assessing learning is not a complete indicator of a curriculum's effectiveness. When self-assessment is similar or on par with expert assessment, it indicates that the trainees understand the intended outcomes expected from their performance, their understanding of the procedural and holistic requirements of the tasks leaving room for improvement of skills where necessary. Implementing assessment protocols thus bears great importance and should be tailored to how a curriculum is structured to make it effective. We explored this in Chapter 4 as discussed above and further made amendments to assessment protocols and training participants in using those assessment tools.

In addition to the use of validated tools and methods for assessment, the role of technical skills in MAS and the dependence on equipment has to be considered. Using VR and augmented reality simulators has the benefit of objectifying the performance of skills and provides feedback that can be used to improve the progression of skills. Whilst this is crucial and has been proven to quantify skills in a training setting, the use of similar objective tools in the OR where performance concerning patients is more relevant and valuable; there is little evidence of practice and advances in this domain. One element of effective training is the use of tools to replicate real life settings that can be achieved by using the same equipment. Another element would be the use of similar assessment protocols. The use of standardized assessment protocols that include forms and assessment by trainers has been the mainstay for use in both skills training and assessment in the OR. However, the use of objective assessment tools that include tracking of instruments without any additional equipment has not been observed in this research. In Chapters 6 and 7, we explore the use of software-based assessment in the OR for objective assessment of MAS performance. The resulting study evaluates the use of motion tracking software and the correlation of data with performance indices. Training in MAS is not linear and surgeons constantly retrain and reevaluate skills keeping up with constant advances in techniques and tools. The technical skills and assessment criteria learned in training labs would benefit the trainees if they can be assessed in a similar fashion in the OR wherein, they can reference their performance to that in real-life procedures and vice-versa.

MODALITIES, TECHNOLOGIES AND METHODS



Technology in medicine has been evolving at an unprecedented and compounding pace in the 21st century. [21–23] This is even more pronounced in MAS that is strongly dependent on technology. [24–27] Similarly, there are several training tools available as summarized in Chapter 2 to surgical trainees and surgical practitioners to keep up with changing practices and technology. These tools enable safe practice outside the operation theatre and enable repetitive practice. Furthermore, these tools enable for objective evaluation of performance and enable self-learning possibilities. This has been proven to account for a more substantial and long-term retention of knowledge and further the development of technical skills based on reflective practice. This concept of learning has been proven effective in different fields of learning ranging from aviation training to even learning to play video games. [28–31] As is evidenced in this thesis, it is clear that technology plays a crucial and multifaceted role in multiple aspects of training in MAS.

In addition, novel modalities in combination with advanced technology are increasingly playing a crucial role in training MAS skills. The training of MAS skills for a long period has been focused on training the surgeon alone. However, as several studies point out, surgeons do not operate in a bubble and thus should not be trained in one. [32] There are several aspects to this statement that bear relevance to the current trends in MAS training. The importance of team training, that resembles the surgical team in the OR and the importance of procedural disruptions that resemble true to life scenarios. In Chapter 10, the study aimed to gather the effects of surgical procedural disruptions and the effects they have on a surgeon's attentiveness. This was measured by a combination of observational, physiological markers and post-operative interviews and the interrelation between them. The physiological markers that included heat flux, metabolic equivalent of tasks and galvanic skin response provided a valuable insight into understanding the effects disruptions have on a surgeon's performance. The combination of head mounted VR technology and the existing simulators can be used to recreate the immersive environments to generate the same physiological responses within trainees. This makes for better transition from skills labs to OR and accurate evaluation of skills under realistic scenarios.

In Chapter 11, we explore this concept of using virtual environments using head mounted displays and VR simulators to evaluate the effect of training surgeons under duress lead to greater immersion and engagement with the tasks. Though the outcomes of performance tend to be lower than that of standalone VR simulators, the implications for the performance bears greater semblance to real life performance. Thus, training surgeons using these realistic environments ensures better transfer of skills, accurate assessment and faster adaption to the OR experience. Interestingly, the multi-institutional study yielded different results with trainees expressing better engagement due to the novel nature of the technology but also expressing the need for customization. This to enable better resemblance to regional personnel dynamics, language, standard operating procedures even including music and ambiance.

FUTURE PROSPECTS

The compounding pace of technology and changing trends in how education is shifting from traditional teaching models to a more self-learning model has significant implications to where the future of MAS training is headed. Since the start of this research in 2015, we had to modify and adapt the methodology to changing trends in assessment protocols and training technologies year after year. Even today, a majority of studies included in this thesis are being expanded and developed to keep up with changing trends; a clear indication of the transformational pace brought about by an increasing number of stakeholders including but not limited to software, hardware, psychology, design and educational professionals. The direction and aim of all these changes, however, appear to converge upon an increasing shift to a trainee centered design of training elements with a strong focus on self-learning, personalization and self-evaluation.

In the areas of MAS assessment, there is a gradual yet consistent development in artificial intelligence and machine learning. [33–35] These novel assessment tools - which are either software-based or technology/equipment dependent - are used to analyze procedure specific component tasks and with further data analysis even predict risk of error based on a host of information. The information can be crowd-sourced based on common errors based on patient history and even pinpoint to surgeon-specific risk rate based on historic trends. [36] These trends were the cornerstone for the studies in Chapter 6 and 7 were conducted with the aim of creating thresholds for performance for technical skills in MAS using such automated tools. Major OR equipment manufacturers are developing integrated and connected OR suites that make analysis of surgical procedures that much easier with all the tools required to document and archive surgical videos in place. [37–39] With the current trend of Internet of Things (IoT) in surgical systems and global data analysis we envision a common platform wherein surgeons can compare their surgical performances with those considered expert surgeons in the world and aspire to improve and progress their skills as a result. Furthermore, these IoT systems can provide an element of constant surveillance and recording that could prevent error

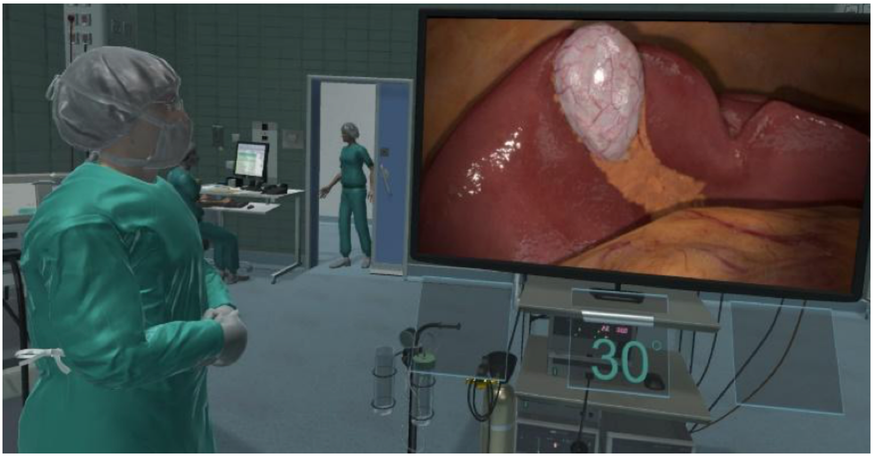


Figure 12.1: An animated OR and characters within a VR simulator

rates and enhance patient safety as a result. On the implementation and curriculum front, there is a fundamental change developing into how training is structured. As it stands, the focus of most training programs tend to focus on some aspects of training be it technical skills or non-technical skills. Recent trends in surgical education design, however, are moving towards a systematic and encompassing model where pathways of procedures are mapped and different stakeholders trained accordingly. [40, 41] These simulation care pathways are procedure specific, region specific and even institution specific. These pathways follow the admission of the patient into hospital, preoperative care, intraoperative and postoperative settings. Herein, different elements of training are used. For instance, in the pre and postoperative settings, mannikins and the team involved are used to simulate that specific setting where critical team training and non-technical skills can be learned. In the intraoperative care, VR simulators with added tools for assistants and surgical nurses are employed to train surgical skills and team training skills. These care pathways ensure a comprehensive training amongst all stakeholders involved towards a common goal of enhanced patient safety and outcomes.

Immersion and personalization within the VR simulator framework are a major part of this research. The conclusions of Chapters 10 and 11 evidenced the need for personalization and its role in immersion and engagement. Current head mounted VR displays employ the use of an animated OR room with generic animated characters as depicted in Figure 12.1. Use of head mounted VR displays with real world 360° videos have been proven to be more immersive than images generated by animation. [42–45]

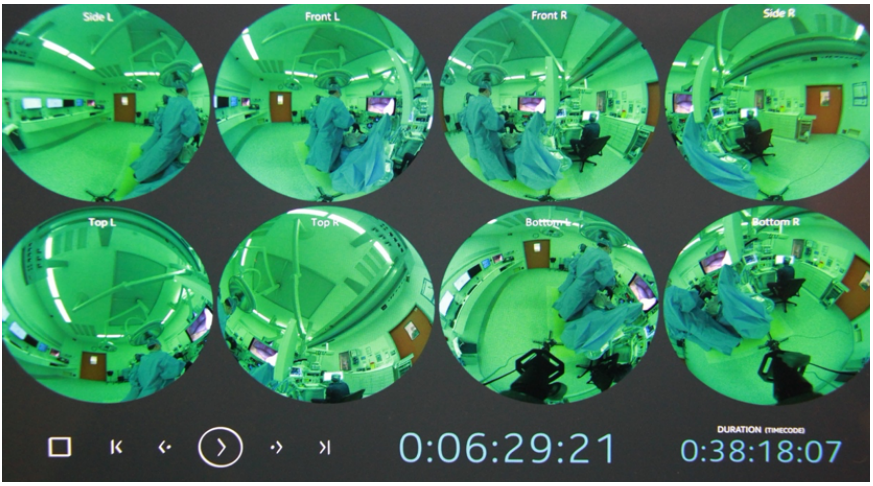


Figure 12.2: Spherical cameras capturing various frames of the OR

Hence, we captured the OR experience during a procedure from the surgeons' point of view using an array of spherical cameras to generate the 360° experience. During the recording, the surgical staff were positioned in places pertaining to the procedure and the dialogue was recorded (Figure 12.2). A sample of the spherical sequence can be seen in the image below. The resulting video is presenting 3D virtual OR-environment made available on YouTube (http://youtu.be/fE9ZOWy06yE) where users can experience the 360° OR environment video. The combination of video, photogrammetry to stitch the videos together and an audio source were used to render the video. Two different possibilities were envisioned during the development of the recording protocols. In the standard case, using game engine technology, the VR simulator output can be imposed in the 360° OR environment video in place of monitors and users get to experience the OR with realistic dialogue and interaction of team. In the other scenario, the video can be projected in a room using an array of projectors wherein the surgical team can train in an immersive environment.

The implications of this extend beyond surgical teams and as such prove to be invaluable and affordable for many institutions where recreating simulated OR experience is not feasible. One of the main challenges in recreating team training environments is the unique and dynamic nature of teams and the challenges in coordinating inter-professional scheduling. Time to talent is another factor to consider where novice healthcare providers face problems in adapting to newer workplace environments and teams. Further research is to be conducted to see if this technology could act as an interim tool to enhance team training outcomes and if the realistic environments in the VR modules could shorten the time to talent which is critical in healthcare delivery.

REFERENCES

[1] A. Darzi and S. Mackay, *Recent advances in minimal access surgery,* BMJ **324**, 31 (2002).

[2] J. C. Rosser, M. Murayama, and N. H. Gabriel, *Minimally invasive surgical training solutions for the twenty-first century,* Surgical Clinics of North America **80**, 1607 (2000).

[3] E. B. Bass, H. A. Pitt, and K. D. Lillemoe, *Cost-effectiveness of laparoscopic cholecystectomy versus open cholecystectomy,* The American Journal of Surgery **165**, 466 (1993).

[4] H. E. Mächler, P. Bergmann, M. Anelli-Monti, D. Dacar, P. Rehak, I. Knez, L. Salaymeh, E. Mahla, and B. Rigler, *Minimally invasive versus conventional aortic valve operations: a prospective study in 120 patients,* The Annals of Thoracic Surgery **67**, 1001 (1999).

[5] R. J. Mannion, A. M. Nowitzke, J. Efendy, and M. J. Wood, *Safety and efficacy of intradural extramedullary spinal tumor removal using a minimally invasive approach,* Operative Neurosurgery **68**, ons208 (2011).

[6] R. A. Steinbrook, *Anaesthesia, minimally invasive surgery and pregnancy,* Best Practice & Research Clinical Anaesthesiology **16**, 131 (2002).

[7] A. Biondi, G. Grosso, A. Mistretta, S. Marventano, C. Toscano, F. Drago, S. Gangi, and F. Basile, *Laparoscopic vs. open approach for colorectal cancer: evolution over time of minimal invasive surgery,* BMC Surgery **13**, S12 (2013).

[8] D. Moon, M. El-Shazly, C. Chang, T. Gianduzzo, and C. Eden, *Laparoscopic pyeloplasty: Evolution of a new gold standard,* Urology **67**, 932 (2006).

[9] M. G. Hartwig and T. A. D'Amico, *Thoracoscopic lobectomy: The gold standard for early-stage lung cancer?* The Annals of Thoracic Surgery **89**, S2098 (2010).

[10] S. J. Symons, P. S. Bhirud, V. Jain, A. S. Shetty, and M. R. Desai, *Laparoscopic pyeloplasty: Our new gold standard,* Journal of Endourology **23**, 463 (2009).

[11] A. Herrmann and R. L. D. Wilde, *Laparoscopic myomectomy—the gold standard,* Gynecology and Minimally Invasive Therapy **3**, 31 (2014).

[12] D. Stefanidis, C. E. Acker, D. Swiderski, B. T. Heniford, and F. L. Greene, *Challenges during the implementation of a laparoscopic skills curriculum in a busy general surgery residency program,* Journal of Surgical Education **65**, 4 (2008).

[13] W. C. Buschemeyer, D. K. Cunningham, and M. J. Edwards, *Surgical training and implementation of emerging surgical technologies,* The American Journal of Surgery **190**, 166 (2005).

[14] J. Sándor, B. Lengyel, T. Haidegger, G. Saftics, G. Papp, Á. Nagy, and G. Wéber, *Minimally invasive surgical technologies: Challenges in education and training,* Asian Journal of Endoscopic Surgery **3**, 101 (2010).

[15] *Curriculum Development for Medical Education: A Six-Step Approach* (Johns Hopkins University Press, 2016).

[16] G. P. Pisano, R. M. Bohmer, and A. C. Edmondson, *Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery,* Management Science **47**, 752 (2001).

[17] F. W. Hafferty, *Beyond curriculum reform,* Academic Medicine **73**, 403 (1998).

[18] T. Shimizu, Y. Tsugawa, Tanoue, R. Konishi, Y. Nishizaki, M. Kishimoto, T. Shiojiri, and Y. Tokuda, *The hospital educational environment and performance of residents in the general medicine in-training examination: a multicenter study in japan,* International Journal of General Medicine , 637 (2013).

[19] R. M. Satava and S. B. Jones, *Virtual environments for medical training and education,* Presence: Teleoperators and Virtual Environments **6**, 139 (1997).

[20] D. Kostons, T. van Gog, and F. Paas, *Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning,* Learning and Instruction **22**, 121 (2012).

[21] A. Weaver and R. Bryce, *Technological advances in medicine: It's personal [guest editors' introduction],* Computer **48**, 21 (2015).

[22] R. A. Atun and D. Sheridan, *Editorial: Innovation in health care: The engine of technological advances,* International Journal of Innovation Management **11**, v (2007).

[23] J. Spiby, *Advances in medical technology over the next 20 years,* Journal of Public Health **10**, 273 (1988).

[24] J. Dankelman, J. J. van den Dobbelsteen, and P. Breedveld, *Current technology on minimally invasive surgery and interventional techniques,* in *2011 2nd International Conference on Instrumentation Control and Automation* (IEEE, 2011).

[25] S. Atallah, M. Albert, and S. Larach, *Transanal minimally invasive surgery: a giant leap forward,* Surgical Endoscopy **24**, 2200 (2010).

[26] M. Tonutti, D. S. Elson, G.-Z. Yang, A. W. Darzi, and M. H. Sodergren, *The role of technology in minimally invasive surgery: state of the art, recent developments and future directions,* Postgraduate Medical Journal **93**, 159 (2016).

[27] J. E. A. Wickham, *Minimally invasive surgery: Future developments,* BMJ **308**, 193 (1994).

[28] E. Verdaasdonk, J. Dankelman, M. Schijven, J. Lange, M. Wentink, and L. Stassen, *Serious gaming and voluntary laparoscopic skills training: A multicenter study,* Minimally Invasive Therapy & Allied Technologies **18**, 232 (2009).

[29] M. Graafland and M. P. Schijven, *A serious game to improve situation awareness in laparoscopic surgery,* in *Games for Health* (Springer Fachmedien Wiesbaden, 2013) pp. 173–182.

[30] L. T. D. Paolis, *Serious game for laparoscopic suturing training,* in *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems* (IEEE, 2012).

[31] M. K. Schlickum, L. Hedman, L. Enochsson, A. Kjellin, and L. Felländer-Tsai, *Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: A prospective randomized study,* World Journal of Surgery **33**, 2360 (2009).

[32] J. R. Pluyter, A.-F. Rutkowski, and J. J. Jakimowicz, *Immersive training: breaking the bubble and measuring the heat,* Surgical Endoscopy **28**, 1545 (2014).

[33] D. A. Hashimoto, G. Rosman, D. Rus, and O. R. Meireles, *Artificial intelligence in surgery,* Annals of Surgery **268**, 70 (2018).

[34] K.-F. Kowalewski, C. R. Garrow, M. W. Schmidt, L. Benner, B. P. Müller-Stich, and F. Nickel, *Sensor-based machine learning for workflow detection and as key to detect expert level in laparoscopic suturing and knot-tying,* Surgical Endoscopy **33**, 3732 (2019).

[35] I. Andras, E. Mazzone, F. W. B. van Leeuwen, G. D. Naeyer, M. N. van Oosterom, S. Beato, T. Buckle, S. O'Sullivan, P. J. van Leeuwen, A. Beulens, N. Crisan, F. D'Hondt, P. Schatteman, H. van Der Poel, P. Dell'Oglio, and A. Mottrie, *Artificial intelligence and robotics: a combination that is changing the operating room,* World Journal of Urology (2019), 10.1007/s00345-019-03037-6.

[36] S. B. Deal, T. S. Lendvay, M. I. Haque, T. Brand, B. Comstock, J. Warren, and A. Alseidi, *Crowd-sourced assessment of technical skills: an opportunity for improvement in the assessment of laparoscopic surgical skills,* The American Journal of Surgery **211**, 398 (2016).

[37] K. M. Irion and P. Novak, *Systems workplace for endoscopic surgery,* Minimally Invasive Therapy & Allied Technologies **9**, 193 (2000).

[38] H. Feussner, *The operating room of the future: A view from Europe,* Surgical Innovation **10**, 149 (2003).

[39] R. Bharathan, R. Aggarwal, and A. Darzi, *Operating room of the future,* Best Practice & Research Clinical Obstetrics & Gynaecology **27**, 311 (2013).

[40] L. Beyer-Berjot, V. Patel, P. Sirimanna, D. A. Hashimoto, S. Berdah, A. Darzi, and R. Aggarwal, *Implementation of a surgical simulation care pathway approach to training in emergency abdominal surgery,* World Journal of Surgery (2019), 10.1007/s00268-019-05242-1.

[41] R. Aggarwal, *Simulation in surgical education,* in *Advancing Surgical Education* (Springer Singapore, 2019) pp. 269–278.

[42] A. Tse, C. Jennett, J. Moore, Z. Watson, J. Rigby, and A. L. Cox, *Was I there?* in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17* (ACM Press, 2017).

[43] S. G. Izard, J. A. J. Méndez, F. J. García-Peñalvo, M. J. López, F. P. Vázquez, and P. Ruisoto, *360° vision applications for medical training,* in *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM 2017* (ACM Press, 2017).

[44] Y. Pulijala, M. Ma, and A. Ayoub, *VR surgery: Interactive virtual reality application for training oral and maxillofacial surgeons using oculus rift and leap motion,* in *Serious Games and Edutainment Applications* (Springer International Publishing, 2017) pp. 187–202.

[45] S. Yoganathan, D. Finch, E. Parkin, and J. Pollard, *360° virtual reality video for the acquisition of knot tying skills: A randomised controlled trial,* International Journal of Surgery **54**, 24 (2018).

CURRICULUM VITAE

Sandeep Ganni was born on November 26th 1987 in Rajahmundry, India. After graduating with an MBBS from GSL Medical College in Rajahmundry in 2011, he moved to Coventry, UK to pursue an MBA in Health Care Management. He graduated with honours from Coventry University in 2012 with a project investigating "Incorporating Lean Principles in Healthcare Management”

His interest in combining healthcare management with simulation education and a meeting with Prof. dr. Jakimowicz in 2014 led to him start a PhD at TU Delft, the Netherlands in 2014. Concurrently acting as the director of the newly formed GSL Smart Lab, Rajahmundry, India, this state of the art centre has grown to deliver over 600 courses every year in South India. In addition to this, he started Compass Healthcare in 2016 as a pharmaceutical chain with a view to adopting new technology in India. Compass has rapidly grown during this period and now employs over 110 staff with multiple pharmacy sites in South India.

Passionate about education, Sandeep delivered a TEDx talk entitled “Medical Training in the Age of Technology” in September 2019 on training in medical education. At the culmination of this project, Sandeep instigated and managed Simulcon India in December 2019: India’s first conference in health care simulation for all healthcare providers with over 1500 delegates. As a result of the success of this conference, Sandeep has helped establish the Society for Healthcare Simulation in India.

Sandeep now splits his time between India and Seattle, USA where he does consulting via Access Health.

Acknowledgements

I would not have been able to complete a project of this magnitude without the help, dedication and support of a large number of people.

First, Prof. dr. Jack Jakimowicz has acted not only as my promotor but as the major driving force for this project, helping me to both complete it on schedule and as excellently as possible.

We first met when Jack visited GSL in India; his commitment to educational excellence and setting and designing and developing standards for curricula was clear from the first moment we met. I really admire his work ethic; his joie de vivre, determination and constant strive for excellence combine brilliantly into something that I both respect and can only hope to exhibit throughout my career! He has certainly inspired me to try to emulate this in my work. I am very grateful for the faith he showed in me, and indeed this project. I will never forget the endless hours of thought and discussion he has dedicated to refining and improving not only the ideas discussed here from their inception but also their eventual presentation in this thesis.

It was a pleasure to work with Prof. dr. ir. Richard Goossens. He showed an immense amount of trust in this project; he helped to steer the shape of the overall project and supported the ideas generation and overall shape of the project. His wealth of experience in Industrial Design was invaluable in focusing the project into a medisign framework and establishing the need for the work presented here.

More than providing the usual familial support, my father, Prof. dr. Bhaskar Rao

Ganni, is a cornerstone for the ideation and implementation of this project. His career has always sought to educate, first founding GSL Medical College which has trained many hundreds of doctors of the future, including myself. I am grateful that his love of adopting the latest research has allowed this project to flourish: he was one of the first to introduce laparoscopic surgery in India and, more recently, has helped me to make the GSL Smart Lab a reality in which much of this work was conducted. His infectious desire to innovate and embrace the latest educational opportunities has always inspired me and thus lead to my choosing this area of study.

Sanne has been an incredible support during this process. Her superhuman work ethic and fastidious attention to detail have helped me develop as both a researcher and a writer. Her strive for excellence has been infectious and her invaluable refinements and suggestions for our manuscripts have been as appreciated as they have been helpful.

Benjamin has always been at the end of the phone throughout this process - providing both emotional and technical advice from his experience writing a PhD.

Magda was with this project from the start and her expertise in motion tracking helped to formulate the direction of this project in terms of automated motion tracking. I remember especially fondly multiple conversations reassuring me that I would 'get there'.

The GSL Smart Lab Team in India have worked tirelessly to help this project - their willingness to adapt to this new way of thinking and their dedication to and trust in the vision and to push themselves is a testament to this. In particular, without Sameer, this project would not have been able to have included India. On top of a busy surgical career, he is a dedicated educator and implemented and collected the data on LSS curricula in India presented here with great success - thank you Sameer.

I would like to thank all of the surgical team at Catharina Hospital for embracing my research so enthusiastically. Their enthusiasm for research and their proactive collection of feedback and videos have made this research much easier. I am especially grateful that they would do so despite a busy schedule and an already complicated working environment. Especially, I would like to thank Gert who was invaluable and unendingly patient in bringing together all the tools and equipment needed, his support and dedication are really appreciated.

Without the expertise of the hundreds of participants, surgeons and surgical interns from both India and Netherlands, as well as their feedback, we would not have been able to adapt this project so well to the learning. Their generously candid and constructive views have made the future of curriculum design more human-centred.

Completing a project of this magnitude involves a huge amount of participants -

from numerous conversations in passing to the generous sharing of expertise over many months. I am extremely grateful to the many people who have contributed to all the aspects of this project.

Family is an invaluable resource and support in any project of this nature. I am indebted to my family for supporting my decision to move halfway across the world to study! Also, I would not have been able to complete the project without the unending support of my friends - both those who I have known for many years back home and those I have met along the journey of this project: thank you.

PUBLICATIONS

PUBLISHED ARTICLES

1. Ganni S, Chmarra MK, Goossens RHM and Jakimowicz JJ. (2017) Self-assessment in laparoscopic surgical skills training: Is it reliable? Surgical endoscopy 31 (6), 2451-2456

2. Ganni S, Botden SMBI, Schaap DP, Verhoeven BH, Goossens RHM and Jakimowicz JJ (2017) "Reflection-Before-Practice" Improves Self-Assessment and End-Performance in Laparoscopic Surgical Skills Training. Journal of Surgical Education 75 (2), 527-533

3. Ganni S, Botden SMBI, Hamilton BF, Bedi AS, Lomanto D, Ganni, BR and Jakimowicz JJ. (2017) Current State of Training and Evaluation of Laparoscopic Surgical Skills. Journal of Surgery Open Access 3(3)

4. Ganni S, Botden SMBI, Chmarra M, Goossens RHM and Jakimowicz JJ. (2018) A Software-based Tool for Video Motion Tracking in the Surgical Skills Assessment Landscape. Surgical Endoscopy 32, 2994-2999

5. Ganni S, Botden SMBI, Chmarra M, Li M, Goossens RHM and Jakimowicz JJ. (2020) Validation of video motion tracking software for assessing performance in laparoscopic cholecystectomy. Journal of Medical Systems 44, 56

6. Ganni, S., Li, M., Botden, S.M.B.I. et al. (2020) Virtual Operating Room Simulation Setup (VORSS) for Procedural Training in Minimally Invasive Surgery – a Pilot Study. Indian Journal of Surgery. doi: 10.1007/s12262-020-02131-z

7. Jakimowicz JJ and Ganni S (2019) Training and Assessment path in MAS: From basic and advanced to specialty procedures. In the book: Recent advances in minimally invasive surgery.

8. Ijgosse WM, Leijte E, Ganni S, Luursema J-M, Francis NK, Jakimowicz JJ and Botden SMBI. (2019) Competency Assessment Tool for Laparoscopic Suturing: Development and Reliability Evaluation. Surgical Endoscopy. doi: 10.1007/s00464-019-07077-2

9. van Houwelingen BCG, Rutkowski A-F, Ganni S, Stepaniak PS and Jakimowicz JJ. (2019) Effects of Surgical Flow Disruptions on a Surgeon's Resources: A Pilot Study. Surgical Endoscopy. doi: 10.1007/s00464-019-07239-2

10. Ganni BR, Nayak SR, Ganni S and Hassan A. (2018) Evaluation of Basic Laparoscopic Surgical Skills Course Programme for Surgery Residents at a Simulation Lab in South India. International Journal of Surgery and Medicine. doi: 10.5455/ijsm.basic-laparoscopic-surgical-skills-course-programme-for-surgery-residents

11. Patchava A, Ganni S, Sudhakar S, Swamy GC, Sudhan SG and Prakash S. (2017) Effectiveness of Stabilization Exercises and Conventional Physiotherapy in Subjects with Knee Osteoarthritis. International Journal of Research in Pharmaceutical Sciences 8 (4) 542-548

ARTICLES IN SUBMISSION

1. Ganni S, Botden SMBI, Goossens RHM and Jakimowicz JJ. Multi-Modal Course in Minimal Access Surgery (MAS) - Achieving Technical and Cognitive Competence – Submitted to Journal of Mininal Access Surgery

2. Ganni S, Botden SMBI, Nayak SR, Ganni BR, Goossens RHM and Jakimoicz JJ. A Comprehensive Three-Year Validation and Impact Study of a Multi-Modal Training Curriculum on Laparoscopic Surgical Skills Training – Submitted to Journal of Surgical Education

3. Althof JF, Huppelschoten AG, Ganni S, Schoot BC, Piek JMJ and Jakimowicz JJ. Face and content validation of the Laparoscopic Hysterectomy Module on Simbionix LAP Mentor virtual reality trainer – Submitted to Journal of Surgical Education

CONFERENCE PROCEEDINGS

1. Ganni S, Nayak SR, Ganni BR and Jakimowicz JJ (2014) Do Simulation Based Curricula for Laparoscopic Training Improve Operating Room Performance? A Preliminary Study – Conference presentation, ELSA, Bali, Indonesia

2. Ganni S and Jakimowicz JJ (2016) Validating training and assessment in Minimal Access Surgery (MAS) – Preliminary results – Conference paper, World Congress for Endoscopic Surgeons, Shanghai, China

3. Li M, Ganni S, Ponten J, Albyrak A, Rutkowski AF and Jakimowicz JJ. (2020) Analysing an Virtual Operating Room (VOR) Surrounding During Laparoscopic Procedural Training – Surgeon's Evaluation on Usability and Evaluation – Conference paper, IEE VR

(01)09312345678907