Doctoral theses at NTNU, 2019:323

Cecilie Våpenstad

Simulation-based Training and Assessment in Minimally Invasive Surgery

Exploration and Validation

NTNU
Norwegian University of
Science and Technology

Cecilie Våpenstad

Doctoral Thesis

Doctoral theses at NTNU, 2019:323

NTNU
Norwegian University of
Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation
and Medical Imaging

ISBN 978-82-326-4242-7 (printed version)
ISBN 978-82-326-4243-4 (electronic version)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

NTNU

Cecilie Våpenstad

Simulation-based Training and Assessment in Minimally Invasive Surgery

Exploration and Validation

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2019

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation
and Medical Imaging

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Circulation
and Medical Imaging

ISBN 978-82-326-4242-7 (printed version)
ISBN 978-82-326-4243-4 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2019:323

Printed by Skipnes Kommunikasjon as

Simulatorbasert trening og bedømming i minimal invasiv kirurgi

Utforskning og Validering

For å sikre lave komplikasjonsrater i minimal invasiv kirurgi er det viktig at kirurgene har de motoriske ferdighetene som skal til for å utføre inngrepet. Simulatorbasert trening gir muligheter for å øve på motoriske ferdigheter utenfor operasjonssalen uten risiko for pasienten. Videre kan simulatorer brukes til å teste og dermed kvalitetssikre kirurgiske ferdigheter. Bruk og utvikling av simulatorer har økt de siste ti-årene blant annet på grunn av overgangen fra åpne teknikker til mer komplekse minimal invasive teknikker, men også på grunn av økt fokus blant sykehusledelse og den generelle befolkning for å få til opplæring uten risiko for pasientene og en kvalitetssikring av kirurgiske ferdigheter. Potensialet til simulatorbasert trening og bedømming er langt fra fullt utnyttet i dag. For å øke bruken av simulatorbasert trening og bedømming er det viktig med videre utvikling, utforskning og validering, i tillegg til politisk og organisatorisk velvilje. Denne doktorgraden er et bidrag i arbeidet med å utforske og validere simulatorbasert trening og bedømming.

Viktige spørsmål som kommer opp ved bruk av simulatorbasert trening og bedømming er om trening på simulatorer betyr økte kliniske ferdigheter, og om simulatorer er i stand til å påvise (teste) kliniske ferdigheter. Doktorgraden er basert på fem studier som så på om ferdigheter tilegnet på simulator ble overført til kliniske ferdigheter (artikkel 4), om ferdigheter målt på simulator tilsvarte kliniske ferdigheter (artikkel 3 og 5), og hva finnes av simulatorbasert prosedyretrening innenfor laparoskopi, endovaskulær kirurgi og fleksibel gastroenterologiske endoskopi (artikkel 2), og hvordan kirurger opplever simulert taktil tilbakeføring (artikkel 1).

I den første artikkelen utforsket vi kirurgers subjektive opplevelse av taktil tilbakeføring ved å be tjue kirurger prøve simulatorer med og en uten i et blindet krysset oppsett. Kirurgene mente at taktil tilbakeføring var viktig, men at de foretrakk simulatoren som ikke simulerte taktil tilbakeføring, hvor de opplevde å få bedre score. I den andre artikkelen så vi på datasimulatorer som simulerte hele eller deler av laparoskopiske, endovaskulære eller fleksible gastroenterologiske endoskopi inngrep. Vi sammenlignet hva produsentene tilbød (online spørreskjema) med hva vitenskapelige studier sa om dem (litteraturstudie). De fem produsentene som deltok tilbød 78 ulike prosedyreoppgaver, hvor av kun 17 av dem ble funnet igjen i litteraturstudien. I tredje og femte artikkel undersøkte vi om ferdigheter som ble målt på simulatorene (artikkel 3 datasimulatoren LapSim®, artikkel 5 bokssimulatoren Simball® box) korresponderte med kliniske ferdigheter. I begge studiene sammenlignet vi kirurger med ulike erfaringsnivå opp mot hverandre. I den fjerde artikkelen så vi på om ferdigheter som kandidatene tilegnet seg på en datasimulator også var nyttige under en klinisk setting. Seksten siste års medisinstudenter (den eksperimentelle gruppen) øvde på datasimulatoren LapSim® til de nådde gitte nivåer på simulatoren. Så utførte de og en kontroll gruppe (fjorten siste års medisinstudenter) en galleoperasjon på en grisemodell. To erfarne kirurger, blindet for gruppetilhørighet, evaluerte videoene av operasjonen. Deltagerne i kontroll gruppen fikk signifikant bedre score i tre av fire kategorier.

Kandidat: *Cecilie Våpenstad*

Institutt for sirkulasjon og bildeveiledning, Fakultet for medisin og helsevitenskap, NTNU.

Veiledere: *Petter Aadahl, Toril Nagelhus Hernes, Ronald Mårvik*

Finansieringskilde: *Samarbeidsorganet Helse Midt-Norge RHF og NTNU*

Ovennevnte avhandling er funnet verdig til å forsvares offentlig for graden ph.d. i medisinsk teknologi. Disputas finner sted på Øya Helsehus, auditorium ØHA1,
onsdag 13. november, kl. 12.15.

Summary

Changes in surgical techniques and public attention to surgeons' competencies in general, created a growth in simulation-based training in the late 90's. The introduction of minimally invasive surgery, with its distinct set of techniques, gave a breeding ground to new ideas for acquisition of surgical skills outside of the operating room thus reducing the reliance on patients for surgical training. In parallel, an increase in adverse events following the introduction of laparoscopy and landmark cases revealing lack of external control, triggered a desire from healthcare management and other stakeholders to assure, through objective assessment, that surgeons have the skills and competencies necessary to perform surgery safely. In addition, organizational changes due to increased pressures on improved efficiency and reduced working hours, have altered the premises for the traditional apprenticeship model. In this context, the development of simulation-based training that reproduce vital parts of the surgical reality appeared as a promising adjunct to the apprenticeship model. Competency-based education, using simulators to assess surgical skills could fulfil patients' and healthcare managements' demand for greater accountability. However, simulation-based training and assessment require simulators that have enough support of validity for the intended use. Simulation-based training and assessment raises questions of whether training on simulators transfer to improved clinical performance and whether documented simulator competence equal clinical competence. The main objectives of this thesis were to contribute to the inquiry of these questions.

In Paper 1 we explored operators' appraisal of haptic feedback devices on a virtual reality (VR) simulator for laparoscopy, and investigated whether performance scores on simulators with and without haptic feedback differed. Twenty surgeons performed two tasks using two handles with and without haptic feedback in a blinded randomized cross-over set-up. Seventy-nine percent of the surgeons answered that haptic feedback is important, and eighty-five percent of the surgeons said they achieved better performance scores with handles without haptic feedback.

In Paper 2 we explored procedural VR simulation in laparoscopy, endovascular surgery and flexible gastrointestinal endoscopy, through a literature review and an online questionnaire answered by simulator companies. The five simulator companies that answered had 78 procedural tasks, where only 17 of them were found in the literature review having been part of a validation study. We found that hardware-software combinations were sufficiently described in only 12 out of 116 retrieved articles. A large number of procedural tasks were available for training and further were in the pipeline, but most of them were still not part of studies investigating their validity.

Paper 3 and 5 investigated whether documented simulator competence correspond with clinical competence for two different simulators: a VR simulator and a box trainer. Surgeons with different levels of experience tested two tasks on the VR simulator with haptics and four tasks on the box trainer with motion tracking. The tasks on the VR simulator and two of the tasks on the box trainer were not able to distinguish between levels of experience. This was probably due to limited physical resemblance of the haptic feedback on the VR simulator or the surgical space in the box trainer. The precision cutting and suture task was able to distinguish between the different groups; while the precision cutting task distinguished between novices and the two other groups, the suture task distinguished between experts and the two other groups, showing that the timing of an assessment task is important.

Paper 4 investigated transfer of skills from a VR simulator with haptics to a clinical setting. An experimental group (N=16) trained on the simulator until they attained predefined score levels on the simulator, before they and a control group (N=14) performed a cholecystectomy on a porcine organ model in a box. Their videos were rated by two expert surgeons blinded to training status. The control group got significantly better scores on three out of four categories: “depth perception”, “bimanual dexterity” and “efficiency”. There was no difference between the groups for the fourth category: “tissue handling”. We believe additional friction in the haptic handles resulted in a negative training effect.

In conclusion, this thesis has explored and validated simulation-based training and assessment in minimally invasive surgery. It was found that several VR simulated procedures are available on the market for training, though only a limited number have been part of a validation study. The validation studies often inadequately described the handles that were used. We found that training on haptic feedback with unrealistic friction resulted in a negative training effect. Presented in three of the papers, we found that haptic feedback is an essential part of VR simulators, and a feature that developers are struggling to make realistic enough. Further it was observed that assessment tasks should be adapted to level of training. Simulation-based training and assessment has unleashed potential both due to political and organizational aspects on one hand and potential technological improvements and educational evaluation on the other.

Acknowledgments

Fulfilling a Ph.D. is not done without help and support. I want to thank professor Toril Nagelhus Hernes for giving me the opportunity to start at the medical technology research group at SINTEF, and then for encouraging me and supporting me as a supervisor throughout my Ph.D. I am also very grateful to my other supervisors, professor Peter Aadahl and professor Ronald Mårvik, for their support. I thank Ronald Mårvik and the rest of the team at the National Advisory board for Advanced Laparoscopic Surgery (NSALK) Gjermund Johnsen, Kirsten Rønning and Hilde-Merete Klungerbo, for introducing me to laparoscopic surgery and welcoming me to the skills lab. I especially thank Petter Aadahl for helping me putting together this Ph.D. I have learned a lot through our discussions on surgical education and the role of simulation.

Further, I want to thank my co-authors, of whom all have contributed to the fulfilment of the five papers that form part of this thesis. Erlend Fagertun Hofstad, nevertheless, deserves a special remark for his continuous support and insight into skills training and assessment.

I also want to thank my colleagues at SINTEF, especially those at the group of medical technology. I am proud to be part of the group and appreciate the support that you have given. Again, a special thanks to my manager, Trond Størseth, for helping and supporting me through the final months of this Ph.D.

Further, I wish to thank the Liaison Committee between the Central Norway Health Authority and NTNU for the funding that allowed me to conduct this research. I am also grateful to all those who volunteered in the studies that form part of this thesis: a total of 145 last-year medical students, surgical trainees and experienced surgeons. I am also grateful for what I have learned from surgeons, surgical trainees and researchers when e.g. being an instructor at the skills lab at NSALK or when visiting the university hospital in Grenoble and the technological environment around it.

Finally, I wish to thank my family. Olivier, my husband, I thank you for helping me with everything from scientific questions to support during ups and downs. I would not have made it without you. And our sons: Emerik, Emilian and Emrys, I thank you for your support and for just being there, reminding me about the most important things in life.

Contents

Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AL	Angular Length
AV	Average Velocity
BD	Bimanual Dexterity
DP	Depth Perception
GOALS	Global Operative Assessment of Laparoscopic Skills
HW	Hardware
IDLE	Idle Percentage
NSALK	Norwegian National Advisory Unit on Advanced Laparoscopic Surgery
NoS	Number of Submovements
MIS	Minimally Invasive Surgery
MS	Motion Smoothness
OR	Operating Room
OSATS	Objective Structured Assessment of Technical Skills
PL	Path Length
RO	Response Orientation
STL	Stereolithography
SW	Software
VR	Virtual Reality

List of Figures

List of Papers

Papers included in the thesis:

1. Våpenstad C, Hofstad EF, Langø T, Mårvik R, Chmarra MK (2013) Perceiving haptic feedback in virtual reality simulators. Surgical Endoscopy 27(7), 2391-2397.
2. Våpenstad C, Buzink SN (2013) Procedural virtual reality simulation in minimally invasive surgery. Surgical Endoscopy 27(2), 364-377.
3. Våpenstad C, Hofstad EF, Bø LE, Chmarra MK, Kuhry E, Johnsen G, Mårvik R, Langø T (2013) Limitations of haptic feedback devices on construct validity of the LapSim® virtual reality simulator. Surgical Endoscopy 27(4), 1386-1396.
4. Våpenstad C, Hofstad EF, Bø LE, Kuhry E, Johnsen G, Mårvik R, Langø T, Hernes TN (2017) Lack of transfer of skills after virtual reality simulator training with haptic feedback. Minimally Invasive Therapy & Allied Technologies, 26(6), 346-354.
5. Våpenstad C, Hofstad EF, Bernstein TE, Aadahl P, Johnsen G, Mårvik R (2019) Optimal Timing of Assessment Tasks Depending on Experience Level of Surgical Trainees - Minimally Invasive Therapy & Allied Technologies, 1-9.

The papers are attached after the references.

My contribution to the papers:

Paper 1: major role in design of study, data collection, analysis and writing of the paper.

Paper 2: major role in design of study, data collection and analysis of online questionnaire, main role in writing of the paper, minor role in the literature review.

Paper 3: major role in design of study, data collection, analysis and writing of the paper.

Paper 4: major role in design of study, analysis and writing of the paper, main role in data collection.

Paper 5: major role in design of study, analysis and writing of the paper, main role in data collection.

Chapter 1 – Introduction

This thesis is about exploring and validating simulation-based minimally invasive surgery training and assessment. The introduction will take the reader through a short introduction of the history of minimally invasive surgery and surgical education, and how changes in surgical techniques and outside attention to surgery in general, created a growth in simulation-based training and assessment. A few learning theories are presented to support the motivation for simulation-based training from a psychomotor learning theory perspective. Furthermore, surgical simulators and assessment tools are presented, before the introduction ends with a chapter on validation of simulation-based surgical training.

The development of minimally invasive surgery

Surgery is by definition invasive and requires incisions into the body. Surgical techniques that requires large incisions are referred to as open surgery whereas minimally invasive surgery (MIS) encompasses surgical techniques with smaller incision sites. A larger incision means a larger wound, more pain and more time to heal. To limit harm, surgeons have always strived for MIS, but they have been limited by technological means. During the 1990s the surgical disciplines experienced a revolution with the introduction of several new technologies enabling MIS (Kelley, 2008). Today minimally invasive techniques are the preferred alternative for many surgical indications (Behrendt, 2017; Fuchs, 2002; Robinson, 2004).

Pioneering work within MIS were e.g. the studies of the surgeon Kelling in and around 1901 and the internist Jacobaeus around 1910. Kelling worked on stomach insufflation (pneumoperitoneum) enabling working space in a closed abdomen. Jacobaeus performed 97 laparoscopies from 1910-1912 using a trocar with a trap-valve to inflate and inspect the peritoneum mainly to evacuate ascites (Himal, 2002; Kelley, 2008; Litynski, 1997). But it was the technological advances, like laparoscopes with integrated mini-video cameras and reliable endovascular materials, during the late 80’s and the 90’s that made minimally invasive surgery feasible and a turning point in surgical history (Kelley, 2008; Richling, 2006). The breakthrough was the result of innovative surgeons and radiologists who saw how to apply technological advances to the surgical field. Nikolay Volodos performed the first endovascular stent-graft surgery in 1985 (Volodos, 2015) and Philippe Mouret performed the first laparoscopic cholecystectomy in 1987 (Nezhat, 2003; Perissat, 1999). These new techniques meant new skills to master, creating opportunities for the young and creative, and hegemonies to defend for the masters (Perissat, 1999).

The advantages of laparoscopy were soon shadowed by an increase in incidents and accidents (Cuschieri, 1995; Perissat, 1999). The increase in adverse events were used by the conservatives to argue against the new techniques, and the pioneers and the promoters of the new techniques understood that something had to be done. Techniques had to be improved and standardized, but most importantly the new techniques had an important initial learning process that had to be proceeded before trying them out on patients. From this, initiatives to change surgical education arose (Perissat, 1999).

Minimally invasive techniques are as numerous and various as open techniques, but they have some common characteristics and psychomotor challenges (Figure 1). MIS involves no or limited direct palpation, tissue is operated on through instruments that are typically long and stiff (laparoscopy) or long and flexible (endovascular techniques), and the procedures are image-guided, i.e. the tip of the instruments are not directly visible but can be seen on a screen. Many minimally invasive techniques are seen as high-tech procedures involving high-tech equipment that need to be mastered. One might argue that the introduction of MIS raised the complexity of the basic skills that a surgical newcomer had to acquire before being able to operate in the operating room (Berguer, 2001; Subramonian, 2004) (Figure 1). As a consequence, increased awareness regarding ways of training outside of the operating room was born, suggesting the use of simulators (Perissat, 1999).

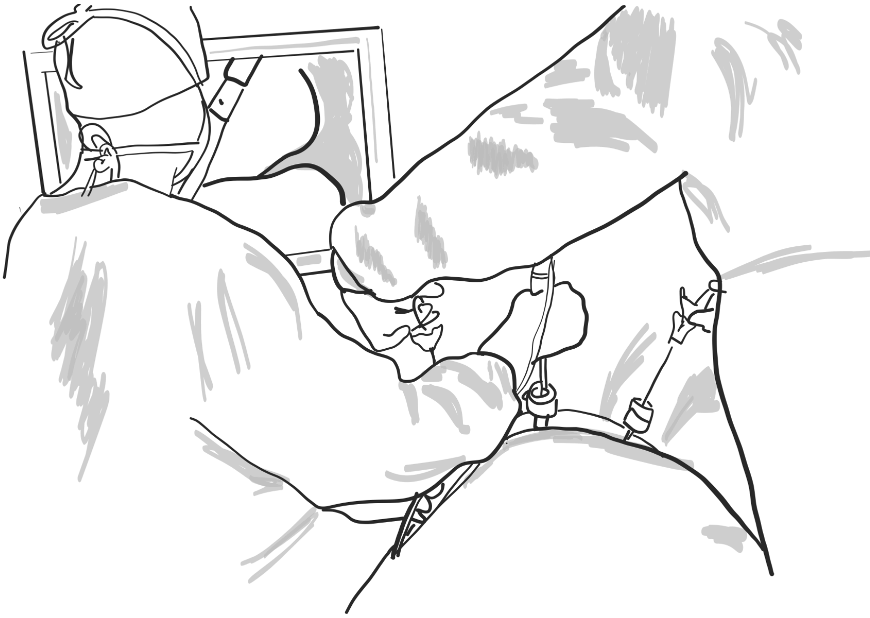

Figure 1: During laparoscopy the surgeon encounters different psychomotor challenges such as the projection of the 3D surgical scene onto a 2D screen by the laparoscope, challenging hand-eye coordination due to e.g. fulcrum effect, limited palpation and challenges related to force transmission. courtesy: Cecilie Våpenstad

Surgical education and learning theories

Surgical education

Surgical education has been based on the apprenticeship model for centuries (Fry, 2011; Hamdorf, 2000). Skills are transferred from master to apprentice, where Halsted is often named the father of the apprenticeship model and the saying "see one, do one, teach one" is used to characterize it (Gorman, 2000; Polavarapu, 2013; Wanzel, 2002). In apprenticeship models or work-based learning, it is not only skills that are taught or learnt, but identities, vocabulary and ways of thinking; not only from the masters, although they are the principal role models, but from the whole working environment, including nurses, other health professionals, engineers, vendors, management and patients (Lave, 1991a; Prentice, 2012).

Halsted's principals of surgical training were based on immersive and repetitive opportunities for surgical care under the supervision of skilled surgical teachers or masters, together with an understanding of the scientific platform of surgical techniques and diseases. The surgical masters would let the residents acquire skills with increasing complexity, giving them gradually enhanced responsibility and independence (Polavarapu, 2013; Prentice, 2012). An apprenticeship model in a well-functioning department or community is robust, as it is not dependent on one person, but rather the whole department or community, which adapts to the needs of the apprentice as the transfer of skills and ways of working are gradual (Lave, 1991a). Surgical departments have governed the education of residents for decades with minimal accountability and external control, and were trusted both by the medical profession and the public (Fry, 2011; Katz, 1999).

An apprenticeship model works well as long as there are no major changes in techniques or ways of working. The changes that were seen during the 90's with the introduction of MIS were obviously major. Although many aspects stayed the same, the former masters were often no masters of the new techniques. MIS required a whole new set of skills, skills that were often counter-intuitive. Adverse events, such as bleedings in MIS, are more difficult to control as ruptures cannot be stopped with direct palpation but only through instruments. The risks and difficulties of these new techniques were at first not fully taken into account, and eager pioneers could try them out on patients with limited external and internal control (Meyers, 1991).

In parallel, patients and healthcare management became increasingly aware of the lack of external control in general, that had resulted in a series of landmark episodes triggering concerns about patient safety (Fry, 2011). One such episode was the Bristol infant heart surgeons' case in the 90's, where it was found that there were surgeons who continued to operate despite having higher mortality and complications rates than fellow surgeons (Smith, 1998; Treasure, 1998). One of the surgeons told the public inquiry that he was on a *"learning curve"* saying that *"Whenever you start any new operation you are bound to have, unfortunately, high mortality"* (Dyer, 1999, p. 1456). A profound change in the relationship between the public, health care management and the profession was seen, resulting in increased attention to surgical training, assessment of surgical skills and sound educational design (Fry, 2011; Smith, 1998). Terms like competency-based, proficiency-based or outcome-based training or education were introduced (Aggarwal, 2006; Gruppen, 2012;

Stefanidis, 2019; van Dongen, 2011). They emphasize the establishment of observable and measurable performance metrics that learners must attain to be regarded as competent (Thinggaard, 2016). Performance metrics can be assessed using e.g. simulators (Stefanidis, 2019), where the defined measurable competencies reflect expectations from internal and external stakeholders (Gruppen, 2012).

Learning theories

In practice, the apprenticeship model in surgical education encompasses the surgical master, the apprentice and the working environment. From an academic and socioeconomic perspective surgical education involves fields such as medicine, education, social sciences, engineering, computer science and healthcare management. Surgeons role in surgical education have been the practical transfer of skills and ways of working through work-based learning, but have also consisted of describing and standardizing techniques, creating step-wise ways of performing procedures, and creating professional vision (Goodwin, 1994), i.e. terms and language to describe and help understand aspects crucial to operating. The role of educators and social scientists have been to understand surgical education and how skills and ways of working are acquired and embodied without doing harm to patients as training objects, and how the process of learning can be as efficient as possible. Engineers and computer scientists have contributed to technology enhanced learning through tools such as simulators.

Surgical expertise is extensive and ranges from teamwork, clinical decision making to technical skills (ACGME, 2019; RACS, 2019; RCS, 2019). As an example The Royal Australasian College of Surgeons has identified nine core competencies that a surgeon need to acquire: *collaboration and teamwork, communication, health advocacy, judgement – clinical decision making, management and leadership, medical expertise, professionalism and ethics, scholarship and teaching, and technical expertise* (RACS, 2019). Needless to say, there is no single training activity where the surgical trainee will acquire them all. Each one of them are complex. With technical skills as an example, just defining (technical) skills has been debated numerous times (Adams, 1987). According to Adams (Adams, 1987, p. 42) a skill is 1) a behavior that is complex , 2) it is learned and *"a scientific understanding of it must* *be concerned with all grades of it"* from perfection to limited mastery and 3) it involves combinations of cognitive, perceptual and psychomotor processes, where the psychomotor component must be present. Research on skills have dwelt with questions related to what the components of nature or nurture are, and how do we as humans learn and retain a skill through distributed versus massed training. Do we attain plateaus or does increasing mastery follow a linear curve with time and effort as input? Do skills transfer from one setting to another? and further how do the conscious – non-conscious components of learning relate (Adams, 1987; Fitts, 1967).

Learning theories are a vast topic, and theories have arisen in non-medical fields and have been adapted to the medical domain, and vice versa. Many theories overlap but differ in use of perspective and focus on different aspects of surgical education (Fry, 2011; Rashid, 2017; Sadideen, 2012b).

Fitts and Posner's three stages of learning

Fitts and Posner created a model on motor skills learning following three phases, focusing on cognitive capacity when learning technical skills (Fitts, 1967). During the first phase, the cognitive phase, the trainee observes and understands the skill. During the second phase, the associative phase, the trainee slowly acquires the skill through trial and failure, and in the third phase, the autonomous phase, the skill is automated liberating cognitive capacity to other aspects when performing it. The three phases evoke some resemblance to the saying "see one, do one, teach one", but the adage "see one, do one, teach one" has been criticized for not taking into account what is needed to pass from one level to the other, or what characterizes the different steps. Time and volume are not a guarantee for taking the trainee through the different phases. In the model of Fitts and Posner, the phases are rather descriptions of level of skills and cognitive capacity available when performing them, and trainees might need different amount of time and effort to attain them. A similar notion was postulated by Adams when he found that *"time-sharing ability"* increased at an advanced stage of training (Adams, 1987). Laparoscopic suturing is often used as an example to make Fitts and Posner's three steps understandable (Sadideen, 2012b). It's a demanding task, and the trainees can spend hours and hours in the skills lab practicing the skill, going through the three phases, liberating cognitive capacitive, which is crucial during the real procedure. Fitts and Posner's three steps theory is a main driver for moving basic skills training out of the operating room into the skills lab, where skills can be practiced without any risk for the patient (Wanzel, 2002). Another important aspect of the theory is that if a skill is so difficult, that it needs the trainee's full attention, then the trainee probably doesn't have that full attention when being a newcomer in the operating room. Suggesting that the trainee might have better learning outcomes when going through the three phases of a complex skill in a less challenging environment, like the skills lab, before entering the operating room.

The theories of closed-loop and deliberate practice

Another motor learning theory is the closed-loop theory introduced by Adams (Adams, 1987). An important aspect of this theory is that the growth in the capability of detecting and correcting errors is key to the learning process. Feedback from response is used to detect error, the errors are corrected and feedback from response is evaluated again and again (Adams, 1987). Perception in the process of error detection is at the heart of the theory, creating what Adams called the *"perceptual trace" (Adams, 1987)*. Similarly, with "*deliberate practice"*, introduced by Ericcson, is the focus on creating *"mental representations"* of what is superior performance, and then to work towards attaining them (Ericsson, 2004; Ericsson, 2008; Ericsson, 2015). Ericcson uses the example of musicians that create mental representations of how the piece of music they are supposed to rehearse should sound when it is played in front of an audience. When rehearsing, any discrepancy between the mental representation of the aspired and the actual performance guides the musician to improvements (Ericsson, 2015). The capability of creating the mental representation of expert performance is crucial to being able to eventually attain it. Simulators with assessment tools producing feedback that is given to the trainee, can aid in identifying expert performance and error detection, and thus help the trainee achieve expert performance in the sense of closed-loop theory or deliberate practice.

Learning theories are explanatory models of how we learn and acquire skills and knowledge. They emphasize different aspects and factors influencing learning and might serve as arguments for moving parts of skills training outside of the operating room. The introduction of radically different techniques and concerns about patient safety, fueled ideas about new ways of learning surgical skills outside of the operating room reducing the reliance on patients for surgical training. Organizational changes have also altered the premises for the traditional apprenticeship model, where service pressures are challenging the conditions for training, and new structures might limit the contact between the masters and the trainees. Simulators that could reproduce essentials parts of the surgical reality seems a good adjunct to the apprenticeship model. Formal assessment using simulators in competency-based education could also answer patients and healthcare managements demand for greater accountability (Fry, 2011; Goldenberg, 2017; Szasz, 2015).

Surgical simulators, principles and core technologies

During the late 90's technological advances not only opened up possibilities for minimally invasive surgery, but also for technology enhanced learning using e.g. virtual reality simulators (Stefanidis, 2019). The use of animals or home-made models for training or testing of new techniques was not new, but the focus and approach on simulation-based training was different and increasing, both by surgeons themselves but also by the simulator industry and health care management (Fry, 2011; Matthews, 2016; Stefanidis, 2019). Research on simulators and simulation-based training emerged simultaneously with MIS during the late 90's (Satava, 1993; Satava, 2001), with a focus on developing (Lamata, 2006a; Liu, 2003; Prentice, 2012) and validating simulators (Aucar, 2005; Carter, 2005; Gallagher, 2003; Gurusamy, 2009; Nagendran, 2013), and how to best integrate them into surgical education (Gallagher, 2005; Sadideen, 2012a; Wanzel, 2002; Windsor, 2009).

Surgical training models

Training models that simulate surgical procedures are diverse, and ranges from simple models to immersive virtual environments (Lahanas, 2016; Yiannakopoulou, 2015). Presented below are the most common training models: cadavers and animal models, box trainers, three dimensional printed models and virtual reality simulators, together with patient-specific simulation.

Cadavers and Animal models

Cadavers and animal models are still used and are highly appreciated by residents for their resemblance to human tissue (Ganpule, 2015; Prentice, 2012; Van Bruwaene, 2015). Operating on cadavers and animals are although limited due to ethical reasons, and lack of infrastructure (Prentice, 2012). The use of organs from animals (left-overs from the butcher) in basic skills courses can help the resident familiarize with natural tissue. Training set-ups using animal models and cadavers often lack objective assessment and feedback, and necessitate that expert surgeons are present for instructions and feedback.

Box trainers

Box trainers can be defined as training models where the user manipulates physical objects as opposed to virtual reality simulators where the user manipulates virtual objects. The objects and their surroundings can be defined as imitators of the surgical scene, hereafter simply called the surgical scene. In box trainers the surgical scene is often projected on to a screen using a camera, often a web camera. Most box trainers are inexpensive and are based on the reproduction of "simple" surgical tasks with little effort made to reproduce sophisticated surgical scenes (Gurusamy, 2014; Li, 2017). Most box trainers lack assessment tools, but that is changing (Li, 2017). Box trainers with assessment tools are often based on tracking of movements using video tracking or some kind of pattern recognition (Hagelsteen, 2016; Hennessey, 2013; Oropesa, 2011).

An important aspect to keep in mind is the distinction between physical resemblance (fidelity) and functional task alignment (Hamstra, 2014; Wanzel, 2002). Physical resemblance is whether the simulator looks and feels like a surgical environment, and functional task alignment is whether training on it transfer to the operating room, or whether it can measure surgical skills. Box trainers often receive lower scores on physical resemblance when compared to virtual reality simulators, but might achieve high scores on functional task alignment (Guedes, 2019; Munz, 2004). What is important is that the box trainer reproduces essential parts of the surgical scene. Natural haptic feedback has been one advantage of box trainers compared to virtual reality simulators.

Three-dimensional (3D) printing

Originally taken from manufacturing and rapid prototyping, 3D printing has seen new uses within surgery, both as a patient-specific pre-planning tool and for surgical training on anatomically realistic models (Malik, 2015; Mitsouras, 2015; Rengier, 2010; Waran, 2015; Watson, 2014). A 3D printed model can be based on most volumetric imaging datasets and after a segmentation of the region of interest and a transformation to the stereolithography (STL) file format, the model can be printed in different materials (Mitsouras, 2015). A handheld model can be tempting and have advantages compared to virtual 3D visualizations. Disadvantages are cost and environmental footprint of printing, difficult to print in materials that resemble human tissue, and that they lack objective assessment tools. With technological advances 3D printers might become more available, and (patient-specific) 3D printed models might be integrated into box trainers or augmented reality simulators.

Virtual reality simulators

In virtual reality (VR) simulators, the surgical scene is virtual, and the user interacts through a physical user-interface that is more or less immersive (Lahanas, 2016; Lamata, 2006a; Våpenstad, 2013a). The surgical scene can be a procedure simulating the organs operated upon with surroundings, or it can be simple tasks such as pick and place of sticks on poles. VR simulators are generally expensive and contain assessment tools. They consist of fidelity resources with software (SW) and hardware (HW) that runs the simulated environment, teaching resources with instructional aids and assessment tools (HW and SW), and the management resources (SW) that handles the users and their courses, tasks and achievements (Våpenstad, 2013a). Figure 2 shows a VR simulator using the taxonomy that was proposed by Våpenstad et al (2013a).

A VR simulator has a physical interface through which the user can interact with the virtual environment. The physical interface can consist of components such as handles with or without haptic feedback, foot pedals and VR glasses (Huber, 2017; Huber, 2018; Våpenstad, 2013a). A simulator that combines real and virtual objects in a common environment, and where the real and virtual objects are spatially registered is called an augmented reality simulator (Botden, 2007; Lahanas, 2016; Loukas, 2013). One of the challenges with an augmented reality simulator is the tracking of the endoscopic instruments and the subsequent realistic integration of them in the virtual world (Loukas, 2013).

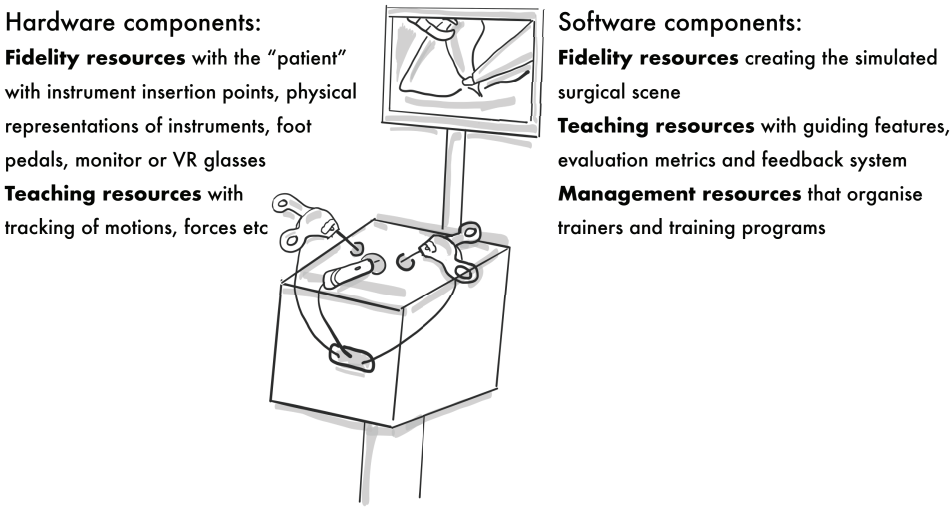

Figure 2: A VR simulator can be divided into hardware interfaces and software components using the taxonomy proposed by Våpenstad et al. (2013a). courtesy: Cecilie Våpenstad

Patient-specific simulation

Patient-specific simulation is a recent development within VR simulators (Willaert, 2012). It has been developed for endovascular procedures such as carotid artery stenting (Cates, 2007; Hislop, 2009; Willaert, 2011) and endovascular abdominal and thoracic aorta aneurysm repair (Davis, 2014; Desender, 2013; Desender, 2017a; Desender, 2017b). In patient-specific simulation or rehearsal, patient-specific data are imported into the simulator, and the operators can perform the procedure on the simulator before they perform it on the patient. Patient-specific rehearsal raises questions towards expectations on operator preparedness before operating on patients.

Haptics – The creation of sensations of touch and proprioception

Acquiring psychomotor skills is about sensing and controlling muscles and joints, and the field of haptics is therefore central to skills learning. In a similar way that vision is about seeing using the senses of sight, haptic is about perception and psychomotor control of objects using the senses of touch and proprioception, where touch is the perception of texture, pressure, vibration, heat and cold, and proprioception or kinesthesia is the perception in muscles and joints of body movement and position (Han, 2016). Haptic technology is the technology that deals with how to reproduce the sensations of haptics: the physical interface, together with SW (Basdogan, 2004; Overtoom, 2018; Ruthenbeck, 2013). Haptic technology is often referred to as haptics or haptic handles in the literature concerning VR simulators (Basdogan, 2004; Coles, 2011; Lamata, 2006a; Våpenstad, 2013c).

In MIS the haptic sensations are tool mediated as the surgeons or interventional radiologists interact with the patients through instruments. Simulating the haptic sensations in MIS means mainly simulating the sensations related to proprioception, and is thus easier to simulate than haptics in open surgery, where the sensation of touch needs to be accounted for. Although easier to simulate than haptics in open surgery, making haptic technology realistic is one of the main challenges in VR simulator technology (Lamata, 2006a; Lamata, 2006b; van der Meijden, 2009; Våpenstad, 2013b; Våpenstad, 2017; Våpenstad, 2013c). The tool-mediated interaction that needs to be emulated ought *"to emulate perfect rigidity when in contact with a virtual rigid object and be mechanically transparent when moving through empty space"* (Våpenstad, 2013c, p. 2392), and to emulate a continuum of haptic feedback for all kinds of forces applied and returned on virtual objects with different (bio)mechanical properties. In practice, VR simulators struggle to achieve haptic technology with sufficient mechanical performance with regards to frequency response, and fidelity with regards to force reproduction and force resolution (El Saddik, 2012; Lamata, 2006a; van der Meijden, 2009). Humans can perceive frequencies around 1 kHz, compared to visual frame rates around 25 Hz, making available computation time for haptic feedback systems limited. VR simulators are available with and without haptic technology, where those with haptic technology are considerably more expensive (Panait, 2009; Salkini, 2010; Thompson, 2011; van der Meijden, 2009). Figure 3 show a possible solution for how haptic feedback can be simulated, another solution can be a pull and push mechanism.

Haptic sensations in MIS are limited compared to open surgery (Overtoom, 2018; Tholey, 2005; van der Meijden, 2009; Våpenstad, 2017; Våpenstad, 2013c; Westebring-van der Putten, 2008), but have been found to be of importance (Chmarra, 2008; Tholey, 2005; van den Dobbelsteen, 2007; Våpenstad, 2017). Although not fully understood (Bholat, 1999; Lamata, 2006b; Picod, 2005; Thompson, 2011), studies have shown that movements such as grasping and pulling are more easily acquired when learned on simulators with realistic haptic feedback (Botden, 2008; Chmarra, 2008; Kim, 2004; Strom, 2006; Zhou, 2012), that unrealistic haptic feedback might influence surgical performance negatively (Våpenstad, 2017), and that the presence or non-presence of haptics influence performance scores on simulators (Buzink, 2010; Cao, 2007; Hagelsteen, 2019; Våpenstad, 2013c) . Despite the fact that haptics influence performance scores, and acquisition and transfer of skills, haptic technology is often poorly described in the literature (Våpenstad, 2013a; Våpenstad, 2013c).

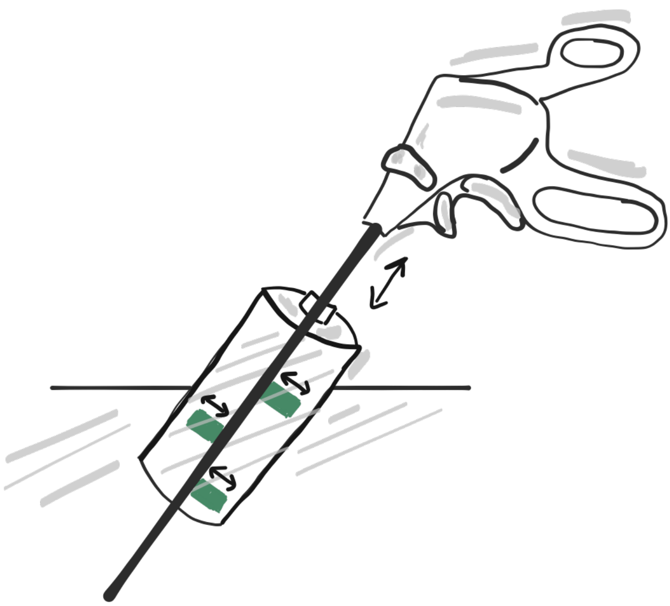

Figure 3: A simplified model of a haptic feedback device. The sensation of haptics is simulated by a set of actuators in the imitated trocar that exert a force on the instrument depending on the virtual object and the force the user applies to it. courtesy: Cecilie Våpenstad

Vision - The creation of the virtual surgical scene

Any simulator will consist of what you see (vision) and what you feel or touch (haptics). The visual system deals with what you see consisting of displays (screens or VR glasses), cameras (for box trainers), graphic cards, cpu's and visualization software. For a box trainer the visual system consists of a camera and the screen the image is projected on. For a VR simulator the visual system is based on simulated 3D models and computer graphics, and is much more complex.

The visual system for a VR simulator is a tradeoff between realism, frequency response and computational power (Ruthenbeck, 2013). A combination of animation and simulation is often used. Inter-active real-time modelling of non-rigid bodies is demanding, where the simulated tissue must respond to the user's movements in a realistic way, indicating frame rates above 25 Hz (Grimm, 2005). That means around 40 ms computation time for collision detection, simulation and visualization (Grimm, 2005) simultaneously with haptic detection and return to the user.

Assessment tools

In competency-based or proficiency-based education, simulators can be used to improve and assess skills and competencies (Bann, 2003; Oropesa, 2011; Stefanidis, 2019; van Hove, 2010). The assessment of surgical competencies can be used during training to give formative and summative feedback improving the training, and to assure that surgical trainees attain pre-defined levels of surgical expertise, either as selection criteria or as a quality assurance (Vassiliou, 2011). Assessment tools can either be automatic, semi-automatic or involve the supervision of surgical experts or alike. Assessment of psychomotor skills can be divided into time measurements, error detection, motion analysis or force analysis.

Time measurements

Time spent on a task is probably the easiest measurable indicator of surgical proficiency. The idea is that a novice spends more time completing a task than an expert, and thus, using less time means being more proficient (Oropesa, 2011). The compelling simplicity of the statement is not always that straightforward. Time can easily be measured, but a task needs to be performed safely and correctly as well, and verifying those aspects are more complicated. Nevertheless, using time as a measure is useful, especially with regards to simple tasks such as pick and place of sticks.

Error detection

Assessment of errors involves the definition of tasks, goals and errors, and can be rather complicated. The easiest, but less practical, would be to manually verify the error, e.g. by verifying that a suture holds and that there is no cleavage. An automatic error detection could be to e.g. verify on a VR simulator that the trainee makes the correct movements or path of movements, does not damage the tissue and do so within a given time frame, creating a good (enough) suture. Automatic error detection that makes sense can be difficult to achieve both on box trainers and VR simulators, as it might seem difficult to include equally good but different ways of resolving a task.

Motion analysis

Tracking motions and motion analysis has been part of VR simulators for a long time (Oropesa, 2011), and it is becoming available also on box trainers (Hagelsteen, 2016; Hennessey, 2013; Li, 2017). The idea is that a surgeon who controls his movements, i.e. has shorter path lengths, will do less harm to a patient. Simulators have therefore been developed so that they can track movements. Data have been presented to the users as formative feedback and often by comparing peers with peers, or educators have created passed-fail levels as part of competency-based education. The idea has often been that shorter time and shorter path length is better, ignoring the complexity of surgical skills (Hofstad, 2017; Oropesa, 2018). In a study investigating motion analysis during cholecystectomies, we found e.g. that novices used their cutting instrument far more, than the experienced surgeons, who used their grasper to move the gallbladder to the cutting instrument (Hofstad, 2017). A simple "shorter path length is better" would yield correct answers for the cutting instrument, but not for the grasper.

Motion-tracking can be based on recording and analyzing electromechanical, optical, mechanical, ultrasound and/or video signals (Oropesa, 2011; Våpenstad, 2013a).

Motion-related metrics can be defined from the position and the orientation of the instruments, where the position can by defined by:

$$\mathrm{r}\ (\mathrm{t}) = \ [\mathrm{x}(\mathrm{t}), \mathrm{y}(\mathrm{t}), \mathrm{z}(\mathrm{t})]_{\mathrm{t}=0}^{\mathrm{T}}$$

and the orientation defined by the three angles:

$$[\alpha(\mathrm{t}), \beta(\mathrm{t}), \gamma(\mathrm{t})]_{\mathrm{t}=0}^{\mathrm{T}}$$

where time (T) is the time from start to completion of the task, measured in seconds, z is the direction of the instrument, x and y indicate the directions in the plane perpendicular to the instrument's axis, and where α and β indicate the rotation around x and y, and γ the rotation around the instrument's axis (Figure 4).

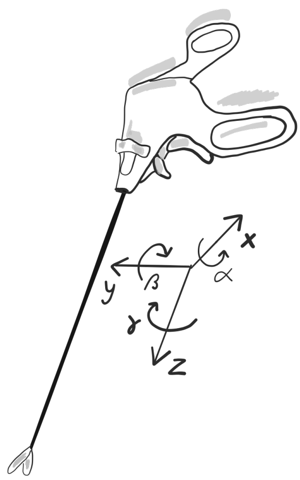

Figure 4: The definition of the position and the orientation with regards to the instrument, courtesy: Cecilie Våpenstad

The following metrics can then be calculated to evaluate the performance of the participants (Hofstad, 2017; Hofstad, 2013; Våpenstad, 2019):

1. *Bimanual dexterity* (BD) is the participant's ability to control and manipulate two instruments at the same time. BD can be found by calculating the correlation between the velocity of the tip of the instruments controlled by the left and the right hand:

$$BD = \frac{\int_0^T (v_{left}(t) - \bar{v}_{left})\big(v_{right}(t) - \bar{v}_{right}\big)dt}{\sqrt{\int_0^T (v_{left}(t) - \bar{v}_{left})^2\, dt \cdot \int_0^T (v_{right}(t) - \bar{v}_{right})^2\, dt}}$$

   where v is the velocity of the instruments and $\bar{v}$ denotes the average velocity over the duration of the task.

2. *Path length* (PL) is the movement of the tip of the instrument, integrated over the duration of the task, measured in meters:

$$PL = \int_0^T \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\ dt$$

3. *Angular length* (AL) is the change in angle of the tip of the instrument in the plane perpendicular to the instrument's axis, integrated over the duration of the task measured in degrees:

$$AL = \int_0^T \sqrt{\left(\frac{d\alpha}{dt}\right)^2 + \left(\frac{d\beta}{dt}\right)^2}\ dt$$

4. *Depth perception* (DP) is calculated by the total distance traveled by the tip of the instrument in the instrument's axis direction, measured in meters:

$$DP = \int_0^T \left|\frac{dz}{dt}\right|\ dt$$

5. *Response orientation* (RO) is a measure of the total amount of instrument rotation around its axis, measured in degrees:

$$RO = \int_0^T \left|\frac{d\gamma}{dt}\right| \, dt$$

6. *Motion smoothness* (MS) is the change in acceleration of the tip of the instrument integrated over the duration of the task and normalized by the duration of the task, measured in m/s^3

$$MS = \sqrt{\frac{1}{2T}\int_0^T \left(\frac{d^3x}{dt^3}\right)^2 + \left(\frac{d^3y}{dt^3}\right)^2 + \left(\frac{d^3z}{dt^3}\right)^2 \, dt}$$

7. *Number of submovements* (NoS) is the number of times a movement of the tip of the instrument contains a velocity peak of at least 10 mm/s.

8. *Average velocity* (AV) is the average velocity of the tip of the instrument, measured in mm/s.

9. *Idle percentage* (IDLE) is the percentage of time that the instrument is moved at a speed below 2 mm/s.

Force analysis

Not applying the appropriate amount of force when operating on human tissue might injure it, causing serious damage (Rodrigues, 2012; Westebring-van der Putten, 2009). Training on VR simulators today, give limited information on whether the correct forces have been applied. Haptic feedback has been found important (Chmarra, 2008; Tholey, 2005; van den Dobbelsteen, 2007; Våpenstad, 2017), and assessment tools that can measure forces and thus give feedback on the use of forces to the user would improve training (Horeman, 2014; Horeman, 2013; Horeman, 2010; Horeman, 2012; Richards, 2000; Rodrigues, 2012). Force measurements and force analysis are rare on today's training devices (Horeman, 2010; Overtoom, 2018; Trejos, 2014).

Force measurement systems can be based on force sensors in the instrument and/or in the environment of the instruments. A force sensor in the instrument measures a force equivalent to the force applied with the instrument on the surroundings, whereas a force sensor in the environment measures a force equivalent of the counterforce of the applied force on the surrounding (Horeman, 2010; Overtoom, 2018). An important part of force analysis is to quantify appropriate levels of forces for given tasks, especially the levels of forces that cause damage (Rodrigues, 2012).

Force-based metrics, similar to motion related metrics can be analysed in multiple ways, where some of the proposed metrics are mean force, peak force, force direction, force volume, force range, force derivates and smoothness of the applied forces (Horeman, 2010; Overtoom, 2018; Rodrigues, 2015; Takayasu, 2018; Trejos, 2014).

Validation of simulation-based surgical training

Validation theory has roots in psychology and pedagogy, and was imported to surgical education when the surgical community saw the need to validate simulators for training and assessment of surgical skills (Carter, 2005; Gallagher, 2003). Although, initially borrowed from the fields of psychology and pedagogy, own interpretations on how to perform and evaluate validation studies within the surgical field emerged and were not updated with methodological advances from psychology and pedagogy (Borgersen, 2018; Cook, 2014; Goldenberg, 2018; Korndorffer, 2010; Sweet, 2010). The currently accepted framework with psychology and pedagogy states that validation is the process of collecting evidence through hypothesis driven studies, investigating the appropriateness, meaningfulness and usefulness of the specific inferences, based on test scores (APA, 2014). Validation is not proven unconditionally, but is based on the collection of evidence of test scores (Korndorffer, 2010; Royal, 2017; Sweet, 2010), i.e. validity can be seen as a continuum onto which weighed and judged evidence cumulates to support an inference (Royal, 2017).

Messick's framework of validity suggests that there are five sources that contribute to the collection of evidence: test content, response process, internal structure, relationship to other variables, and consequences (Borgersen, 2018; Cook, 2006; Goldenberg, 2018). Test content is about relevance of the simulator/test with its intended use, e.g. if the intention of a test is to measure preparedness to operate appendectomies in the OR the requirements of the test content is different than if the intention is to simply say something about surgical psychomotor skills in general (Downing, 2004). Response process evidence represents the data integrity of the test scores, e.g. if some participants have trained far more than the others on the simulator before it is tested, this might bias the interpretation of the results (Goldenberg, 2018). Internal structure evidence relate to reliability and reproducibility of the tested entity (Cook, 2013). Relationship to other variables is the validity evidence source that is mostly performed, and is about the statistical association between assessment scores being tested and a specified theoretical relationship, e.g. variations in simulator scores across number of laparoscopic procedures performed (Cook, 2014), or by comparing assessment scores against a previously validated gold standard (Goldenberg, 2018). The source called consequences is about the consequences of the interpretation of the results, positive or negative, beneficial or harmful, i.e. the results or consequences that might follow of the implementation of a test result. A test investigating consequences can be whether a test that is used to select candidates for residency programs selects well suited candidates. A positive consequence would be if the candidates that are selected are well suited, a negative consequence could be that the test is too strict and good candidates are not allowed to enter, having negative consequences both for the candidates themselves and for society as a whole (Goldenberg, 2018). Higher-stake uses of simulators should demand a larger collection of evidence compared to lower-stakes uses (Korndorffer, 2010; Royal, 2017; Sweet, 2010).

A study of validity, based on test scores, often seeks to investigate parts of a simulator, such as one or more tasks, one or more tests, or one or more motion parameters, within the given context of the study. Presented below are common validation studies as they have been performed by the surgical education community within the old paradigm, highlighting advantages and drawbacks.

Face and content validity or operator's appraisal

The method of face validity is typically to ask experts what they think about the simulator or part of it. It is a subjective evaluation of how they as experts' value it. If the experts are positive towards the simulator or part of it, it is said to have face validity (Gallagher, 2003). Content validity is similar, but questions the content of the simulator (Gallagher, 2003). Assumptions underlying face and content validity are that experts, because they are *"experts"* and have extensive experience and knowledge within the field, have opinions about e.g. a simulator that has value. Face and content validity studies are strictly not validation studies (Korndorffer, 2010) as they do not investigate test scores of the simulators, but investigate subjective opinions. A questionnaire on operator's appraisal is comparable to a face and content validity study, and might be a more correct term.

Construct validity

Construct validity is a measure of whether a test or an assessment tool measures what it claims to measure (Gallagher, 2003; Sweet, 2010). After Messick's framework of validity it can be compared with the source of evidence called relationship to other variables (Borgersen, 2018). The assessment tool or the formative feedback after a task on a surgical simulator is said to show construct validity if it is capable of measuring surgical skills. Construct validity is often measured using the known-groups technique, where it is believed that the groups differ due to specific characteristics. In e.g. laparoscopy, groups are often created based on experience such as number of laparoscopic procedures performed (Fairhurst, 2011; Korndorffer, 2010). It is then assumed that surgeons who have performed more procedures have higher surgical skills than those that have performed less. If the most experienced group get higher scores on a task than the less experienced ones, the study indicates that the task show construct validity. There are differences in levels of surgical skills (Birkmeyer, 2013), also between surgeons with experience. As a consequence, dividing into groups based on experience and not on actual skills level, might have drawbacks.

Studies investigating construct validity on surgical simulators often compare all the metrics of the simulator and then concludes that those where a difference between experience levels were found exhibit construct validity, and the others did not. This approach has been criticized for not having looked into which metrics seemed meaningful before or after the test to explain why that or those metrics distinguished between surgeons with different levels of experience (Korndorffer, 2010). Similarly, a result might be statistically significant, but irrelevant to surgical education or surgical skills training (Sweet, 2010).

Content underrepresentation might also be of concern as important surgical competencies such as situational awareness or time-sharing ability are not tested (Korndorffer, 2010). Differences in scores between novices and experts might simply be due to familiarity to the equipment and to a limit extent a surgical skill. In addition, finding differences between novices without any experience and surgeons with 20 years of experience, might be of limited use, if the intention of the study is to find out if the simulator can be used to decide which residents are ready to operate (Korndorffer, 2010). Such a test would have to investigate skills that occur earlier in a surgeon's career, and at the same time keep in mind weaknesses of content underrepresentation.

As new simulators are developed, performing scientific studies investigating construct validity, or relationship to other variables, stays an ongoing need (Borgersen, 2018; Stefanidis, 2012).

Concurrent validity

Concurrent validity investigates correlation between an assessment tool and other assessment tools, often a so-called gold standard (Tavakol, 2008). Concurrent validation is strictly speaking not a validation, but a correlation between different assessment tools.

Predictive validity or transferability studies

A predictive validity study investigates whether training on a simulator, i.e. the obtainment of good scores on a simulator, predicts better performance (results) in the real setting, i.e. the operating room (Ahlberg, 2007; Seymour, 2002). A group of subjects are usually tested for a certain construct, e.g. performance levels on a simulator, and then those results are compared against results obtained from the real setting, the operating room. Equivalent terms are investigations into transfer of skills or transferability studies. One of the difficulties with studies investigating transfer of skills is confounding variables, another is that it is demanding to follow residents over a long time and into the operating room. The study of Ahlberg et al. (Ahlberg, 2007) was one of the first, and one of few, that followed residents into the operating room investigating the correlation between simulator training, no training and operating room performance.

Chapter 2 - Thesis outline

Objectives

In 2012, nearly twenty years after the onset of interest in simulation-based minimally invasive surgery education, the highest rated questions by the association of surgical education in the USA (Stefanidis, 2012 p. 51) were *"Does training on simulators transfer to improved clinical performance?"* and "*Does documented simulator competence equal clinical competence?"* . The main objectives of this thesis were to contribute to the answer of these questions. In parallel, an objective has been to explore the nature of simulation-based minimally invasive surgery training, with a focus on haptics. As such, the objectives of this thesis were to investigate and to explore:

- Transfer of skills from simulator training to clinical performance (1st Objective).
- The correspondence of documented simulator competence with clinical competence (2nd Objective).
- The characteristics of simulation-based minimally invasive surgery training (3rd Objective).

Paper 4 investigated the 1st objective. Paper 3 and 5 investigated the 2nd objective, and paper 1 and 2 explored the 3rd objective.

The studies of inquiry have been performed in a laparoscopic setting, but are relevant for minimally invasive surgery in general.

The Norwegian National Advisory Unit on Advanced Laparoscopic Surgery (NSALK) – the Infrastructure

The Norwegian National Advisory Unit on Advanced Laparoscopic Surgery (nsalk.org) was founded in 1996 to respond to challenges and opportunities related to the new surgical technique: laparoscopy. Dr. Ronald Mårvik, the head of NSALK, was one of the Norwegian pioneers within laparoscopy. The activity of the center focuses on training and research, and has a skills lab with different training modalities. From 2009 on the center acquired new virtual reality simulators and box trainers. The center wanted to implement them into surgical education in Norway, and to use the simulators to assess surgical skills objectively. This was the backdrop for this thesis.

Today, NSALK has training facilities that can host around 30 participants, they have several box trainers and VR simulators, and has access to an animal lab (Figure 5).

The studies from paper 1, 3 and 4 were performed at the center. The study from paper 5 was performed at the center in addition to the regional hospital in Ålesund.

Figure 5: The skills lab and training facilities at NSALK, courtesy: Cecilie Våpenstad

Summary of papers

This thesis is based on a collection of papers that addresses the objectives.

Paper 1 “Perceiving haptic feedback in virtual reality simulators” explored operators’ appraisal of haptic feedback devices, and investigated whether performance scores on simulators with and without haptic feedback differed. The background of the study was that training of psychomotor skills outside of the operating room using VR simulators was increasing and that haptic sensations influence psychomotor performance. How do laparoscopic surgeons then perceive the emulation of haptic feedback in VR simulators, and does haptic feedback or not affect performance scores? We used two different handles: Xitact IHP with haptic feedback and Xitact ITP without haptic feedback (Mentice AB, Gothenburg, Sweden). Both connected to Lapsim® VR simulators (Surgical Science AB, Gothenburg, Sweden). Twenty surgeons performed two tasks using the two handles in a blinded randomized cross-over set-up, i.e. the surgeons were not told which handles simulated haptic feedback, and the participants were randomized to try one of the two handles first. The surgeons’ perceptions of the handles were explored by asking them, after having tried the handles, to answer 12 questions mainly based on Likert scales and closed-ended questions. We found that 79 % of the surgeons claimed that handles with haptic feedback on VR simulators are important given that they feel realistic (score 4 or 5 out of 5). Eighty-five percent of the participants answered that it is important that a handle that tries to emulate haptic sensations do so in a realistic manner. Eighty-five percent of the surgeons said they got better performance scores on the simulator without haptic feedback. Ninety percent of the participants preferred the handles without haptic feedback, and we believe that this was due to additional friction that was added in the handles that tried to simulate haptic feedback. This made them unrealistic and not mechanically transparent.

Paper 2 “Procedural virtual reality simulation in minimally invasive surgery” explored procedural virtual reality simulation in laparoscopy, endovascular surgery and flexible gastrointestinal endoscopy, through a literature review and analyses of an online questionnaire answered by simulator companies. As basic skills VR simulation was getting more and more common, what could be the role of procedural VR simulation? Could it act as a bridging gap between basic skills VR simulator training and training in the OR? Five simulator companies answered the questionnaire stating that they offered 78 procedural tasks. The literature review was a systematic search in the PUBMED and SCOPUS databases between 1985 and February 2012 using a set of search-words. One hundred and sixteen articles were included and reviewed. Of the 78 procedural tasks, only 17 of them were found in the literature review having been part of a validation study, i.e. 61 of the procedures had to our knowledge not yet been scientifically investigated. Another limitation that we found was that only 12 out of the 116 retrieved articles specified which hardware-software combinations that were used in the studies they described. Most simulators in the study had the possibility to provide both formative and summative feedback, potentially strengthening the learning process. Procedural VR simulation was found to be largely available on the market, but that there is still a need to further explore and validate what exists, and to delineate the level of simulation fidelity that is needed to create simulated environment that show transfer of skills to the surgical setting and that can be used for assessment of surgical skills. The paper also suggested a taxonomy of virtual reality simulators dividing it into *fidelity resources*, *teaching resources* and *management resources*.

Paper 3 "Limitations of haptic feedback devices on construct validity of the LapSim® virtual reality simulator" investigated construct validity or relationship to other variables on the LapSim® VR simulator with Xitact IHP handles with haptic feedback using the known-groups technique comparing the scores from novices, intermediates and experts on two different tasks. All participants, in total 47, performed the tasks *"lifting and grasping"* and *"fine dissection"* 20 times each, trying to pass predefined threshold levels for a set of 19 parameters. Significant difference where the experts met the passing levels more times than the novices was found for one parameter, *"misses on right side"*, whereas the passing level for the *"cutter angular path length"* was met more often by the novices than the experts. At the moment of publication of this study a certain number of other studies had previously tested the two tasks for construct validity: Eleven studies had tried to establish construct validity for the *"lifting and grasping"* task, where three reported construct validity for an overall score, seven reported significant differences between the groups for a number of the parameters tested, and in one study the novices outperformed the experts in 4 of 16 tasks. One study had investigated the *"fine dissection"* task and found significant differences for 3 out of 14 parameters. None of the studies used the same handles as we did. Going through construct validity studies on the LapSim® VR simulator we found that in only 7 of 17 studies, the type of handles used were explicitly described. We argued that the reason, we did not find construct validity or evidence for relationship to other variables, was that the handles had a haptic interface that was not mechanically transparent, i.e. they created unrealistic frictional forces that made all participants novices in front of the simulator.

Paper 4 "Lack of transfer of skills after virtual reality simulator training with haptic feedback" investigated transfer of skills from VR training to the OR comparing an experimental group that trained on the LapSim® VR simulator with Xitact IHP handles with haptic feedback until they reached predefined score levels on the simulator, with a control group that only received theoretical instructions. Videos of both groups removing a gallbladder from a porcine animal model in a box were rated by two experts blinded to training status using the Global Operative Assessment of Laparoscopic Skills (GOALS) scale (Vassiliou, 2005). Medical students in their last years were recruited, in total, 16 in the simulator group and 14 in the control group. The control group achieved significantly better video rating scores than the simulator group on three out of four categories: *"depth perception"*, *"bimanual dexterity"* and *"efficiency"*. The fourth category *"tissue handling"* did not show a significant difference between the two groups. The participants in the simulator group completed the criterion-based training program with a median of 79 trials to pass the five tasks in the program (range 32-162). There was a correlation between a low number of trials to pass and a high video rating score. We also compared, by asking the participants about their computer gaming experience, whether prior computer gaming predicted better scores. We found that participants, from both groups, that played computer games weekly or more frequently (N=7) had significantly better scores on one of the four categories: *"depth perception"*. When comparing with other transfer of skills studies it seems that our negative finding is due to the handles simulating haptic feedback. The handles were not mechanically transparent and introduced additional friction compared to the trocars that are used in the clinic. Our finding that unrealistic haptic feedback has negative training effect, is a finding that ought to be further investigated.

Paper 5 "Optimal Timing of Assessment Tasks Depending on Experience Level of Surgical Trainees" investigated the correspondence of documented simulator competence with clinical competence on a box trainer called Simball® box, i.e. a construct validity study or an investigation of relationship to other variables. The participants, 10 novices (0-10 procedures), 22 intermediates (11-100 procedures) and 16 experts (> 100 procedures), performed four different tasks on the box trainer with motion analysis: *"peg picker"*, "*rope race*", *"precision cutting"* and *"suture"*. Nine different motion parameters in addition to time were analyzed, resulting in a total of eighteen metrics as eight of them were measured separately for each hand. No or limited significant difference were found for the *peg picker* and *rope race*. For the *precision cutting* task 12 parameters showed significant difference between novices and intermediates, 14 between novices and experts and 1 between intermediates and experts. For the *suture* task the corresponding results were 1, 15 and 6. The *precision cutting* and *suture* task, thus, indicate evidence that documented simulator competence correspond with clinical competence for many of the parameters. By comparing between three different experience levels we found that the *precision cutting* task distinguished best between novices and the other two groups, and the *suture* task distinguished best between experts and the other two groups. These results show, that the timing of an assessment task is important, if it is to have value.

Chapter 3 – Discussion and main conclusions

The rise of more complex and technically challenging minimally invasive techniques, as well as the increased scrutiny from the wider public, have motivated research on and development of simulators and simulation-based training and assessment. Today simulation-based training for minimally invasive surgery is a promising adjunct to traditional apprenticeship-based surgical education: training without risk for the patients, tools that can be used to objectively assess surgical skills, and new opportunities to prepare before procedures. Simulators are used here and there (Nicholas, 2019), but maybe not as systematically and to an extent that was hoped for ten or twenty years ago. As Kneebone and Fry writes:

"innovations in simulation, for example, are apt to be taken up enthusiastically and equally readily discarded. Fashion exerts a powerful influence, and there is often a mismatch between the adoption of new approaches and their systematic evaluation. By contrast, educational evaluation moves at a much slower pace than the innovations it is expected to judge." (Fry, 2011, p. 14)

This thesis joins the effort of educational exploration and evaluation of simulation-based training and assessment in minimally invasive surgery.

Simulation-based training initially developed from a wish to train psychomotor skills without reliance on the use of patients, but has a potential to cover also other surgical competencies. The papers presented in this thesis mainly focus on psychomotor skills training and assessment, and whereas some aspects are specific for psychomotor skills training and assessment, others might be generalized to simulation-based training and assessment in general.

Ethical considerations were discussed prior to each of the studies. We regarded participant performance scores on the simulators or animal models to be sensitive information, and approval was sought and gotten at the Norwegian data protection agency for paper 1,3, 4 and 5. Paper 2 did not require an approval. In paper 1 we worked with anonymous participants, during study 3-5 we worked with code names and kept personal information separate from the results. All participants in paper 3-5 gave written informed consent. The porcine livers with intact gall bladders that were used and presented in paper 4 were collected from a local butcher as left overs.

Methodological considerations

Paper 1 was motivated by the wish to explore a rather high-level question: the nature of haptic feedback in general and how it influences training on VR simulators. For practical reasons it was out of scope to compare whether VR simulators with haptics transfer better or worse than VR simulators without haptics to a clinical setting. We ended up asking surgeons to test the same VR simulator equipped with handles with and without haptic feedback in a cross-over set-up, and asking them afterwards how they perceived the simulator. Asking questions gave insight into how the surgeons perceived the two simulators. It can be difficult to investigate exactly how the surgeons perceived the handles: how each surgeon experienced the simulated haptic feedback in their mind? What were his or her subjective experiences? It might turn into a metaphysical question (Prentice, 2012). We interpreted the answers not as subjective experiences or reflections of their minds, but as a question of consistency of interpretation. Our investigation was not whether the surgeons' experiences were identical, but whether they agreed or not on the given answer alternatives.

Likert scales were chosen as a numerical rating scale to derive quantitative measures of perceptions and attitudes, in this case on handles with and without haptic feedback. Likert scales are common and thereby known to most respondents. Likert scales do not force the participant to provide simple yes or no answers, but allow them to grade their degree of agreement or disagreement on a scale, in our case from one till five or one till three. The responses could thus contain neutral or undecided attitudes, and the collected data were easily presented graphically (e.g. boxplots).

Drawbacks with Likert scales are, that in reality perception and attitudes towards a given statement exist on a multi-dimensional continuum, and not in presumably equidistant steps like a 3 or 5-scale Likert scale. The answers are probably influenced by previous questions or which handle they tried first, and some people tend to avoid the extreme options on either side of the scale. Further, questions containing more than one statement might be difficult to analyze as the respondents might have had different attitudes towards the sub-statements and it is difficult to know which statement influenced their answers the most. The answers might also be influenced by a variety of parameters such as age, experience level, previous knowledge and/or prejudices towards the topic. We chose a cross-over design, so that half of the users tried the handles with haptic first and the other half tried those without first, thus minimizing the bias from which handle that was tried first. One question contained a double-statement: *"How important is it that the handle has haptic feedback given that it is realistic"*. In retrospect the *"given that it is realistic"* part of the question could have been removed, making the interpretation of the answers easier. We assumed that the surgeons emphasized the first part of the question when they answered, and that the last part was of minor importance.

When exploring perception and attitudes, as in this study, by asking the surgeons to try out two different handles and then asking them different questions, the role of the investigator is not without importance. I organized the test of the handles and the distribution of the questionnaires, and tried my best to act neutral towards the handles. I did not inform the participants about which handles simulated haptic feedback, and there was only a minor physical difference between the two handles. We also phrased the questionnaire using common terms such as *"sensation of tissue stiffness"* and did not use the terms *"haptic"*, *"tactile"* or *"force feedback"* to avoid misunderstandings or prior prejudices.

We asked the participants about their perception, i.e. with which handles did they perform best. Buzink et al. (2010) found that performance was altered with different handles and that participants achieved higher scores with handles without haptic feedback. We found that 17 out of 20 surgeons said that they performed best with the handles without haptic feedback, thus in accordance with Buzink et al. (2010). This was what we also observed when looking at their scores. In retrospect it would have added additional strength to the paper if we would have calculated the exact numbers as these got registered by the simulators.

All in all, we believe that we explored our lower-level question of how these handles with and without haptic feedback were perceived by the surgeons. Some of the perceptions and attitudes were specific for these two handles, others can probably be generalized to account for handles with and without haptic feedback in general.

Paper 2 was motivated by a wish to explore procedural virtual reality simulation: what exists (the company survey) and what is scientifically known about what exists (the literature review). The online survey contained dichotomous, closed format and open format questions. Using closed format questions as much as possible made the comparison between the different procedural virtual reality simulators easier. The closed format questions were taken out of the taxonomy that we created, e.g. one question was "*What kind of teaching resources, formative feedback and instructional aids does the system offer? - Please check one or more answers*" where there was a list with e.g. "*written dialog boxes*", "*change of colour at collision*", "*audible discomfort (patient noise)*", that the respondent could choose from. The risk with closed format questions are to leave out choices that might be important. Both authors had first-hand experience with virtual reality simulators, and we tried our best to make the list of alternatives as exhaustive as possible.

The survey was sent out to 13 companies and research groups that we found through either prior knowledge, a search on the internet or the literature. Six companies answered of which one company was excluded. None of the research groups answered. The success of a survey depends on who answers. We lacked answers from research groups that might have given insight into the latest developments within procedural VR simulation. Instead the paper focused on what was on the market at that time, and we believe that the most important companies answered.

A literature review is dependent on appropriate search words. We chose relatively broad terms, and analysed a large number of retrieved abstracts, i.e. 1873, by both authors independently, ending up with 116 articles that we selected for a full review. There are certainly articles that we missed, but we believe that the broad terms, together with manually going through and selecting appropriate articles gave a comprehensive and appropriate selection.

Paper 3 and 5 were motivated by the aim to implement competency-based education at hospitals in the region of Mid-Norway and at courses at the Norwegian Advisory Unit for Advanced Laparoscopic Surgery (NSALK). We wanted to investigate a virtual-reality simulator's (paper 3) and a box trainer's (paper 5) ability to measure surgical skills, thus investigating construct validity, or when applying Messick's framework: the relationship to other variables (Borgersen, 2018; Cook, 2013). We used the known-groups technique, which is a common way of investigating construct validity (Våpenstad, 2013b). The assumptions underlying the known-groups technique are that surgeons with similar levels of experience will have similar surgical skills and that surgical skills increases with level of experience. It then follows that if a group of more experienced surgeons scores better on the simulator than a group of less experienced surgeons, then the assessment tool is capable of measuring the surgical skills that the experienced surgeons have and the less experienced surgeons don't have. For practical reasons, the known-groups technique is fairly straight forward: groups of surgeons with different levels of experience need to be gathered and then asked to perform the tasks that are under investigation. The assumptions do seem plausible and the known-groups technique is also widely used. It is common to use number of (laparoscopic) procedures as an indicator of level of experience, as was done in 14 of the 17 studies that we found when comparing construct validity studies on the LapSim® VR simulator (Våpenstad, 2013b). The three other also used the known-groups technique but divided into groups of how difficult the procedures the surgeons performed were (e.g. transplantation surgery versus less complex surgery such as cholecystectomies) (Danila, 2009) or level of medical or surgical education (e.g. medical students versus residents versus faculty) (Duffy, 2005; Woodrum, 2006). However, variations in surgical skills do exist between surgeons having performed similar number of procedures. In paper 3, e.g., did the experts, i.e. those that had performed more than 300 laparoscopic procedures, pass the predefined passing level between 0 and 16 times for the lifting and grasping task. And the variances within each group was larger in the expert group than within the novice and the intermediate group.

Should it be a research topic to develop psychomotor skills tests that can measure actual surgical skills, and to use the results of those when dividing the participants into groups? One major challenge would be that surgical skills are not one single skill, but a large set of more or less complex sub-skills, where one psychomotor skills test cannot test them all, probably not five or ten tests either. This probably leaves us with the known-groups technique as the best option there is today. Using the known-groups technique is not without pitfalls. As we saw in paper 5, choosing appropriate experience levels are important and should be kept in mind. We found that the experience levels that we had were well suited to test the precision cutting task, but that our expert cut-off of 100 procedures seemed too low to investigate where suture skills stop increasing. It is also important to use surgeons with appropriate levels of experience with regards to the intended use of the task or test that is being validated.

The number of test-runs and the use of passed-failed levels might influence the results of a construct validity study. In paper 5 the participants performed the tasks twice, and the results of the second run was recorded and analysed. By using the second run, the participants familiarized with the task in the first run, and misunderstandings of how to perform the tasks were reduced. In paper 3 the participants also performed a test run before they were asked to try to pass predefined passing levels twenty times for each task. We then compared the number of trials the participants passed the tasks (total score and sub-parameters). By comparing threshold values, as in our study, we investigated construct validity of the threshold values and not directly the assessment tool of the simulator. But as the threshold values was based on the assessment tool, the assessment tool was part of the validation study. The overall threshold values seemed strict, especially for the *lifting and grasping* task, as several participants in each group did not succeed in passing them at all. The threshold values were based on the work of Ahlberg et al. (2007) and a pilot study. The work of Ahlberg et al. was based on the same simulator as we used, but with different handles than we had, handles with haptic feedback, but where the haptic feedback functionality was turned off (Ahlberg, 2007). We were at that moment not aware of the fact that handles largely influence performance scores. In retrospect the overall threshold levels could have been less strict, i.e. more adapted to our handles with haptic feedback. When looking at the parameters one by one this was of less importance.

Including a larger number of trials further reduced potential errors of not knowing the simulator, but introduced other biases such as motivation (twenty trials took some time), and that there was a learning effect on the simulator, which could be related to other aspects than surgical skills, such as age.

If there were any major reliability issues on the simulator they might have been discovered during the numerous trials. Performances of the same participant naturally differ from one trial to another, and just comparing the scores participant per participant is not enough to do a proper investigation of reliability or reproducibility. An external measure of the parameters (if possible) would have been needed.

Paper 4 was a study investigating transfer of skills from a VR simulator to a clinical setting. Thirty last year medical students were recruited, and then randomly assigned to either the experimental group that trained on the VR simulator or the control group that did not. Both groups performed a cholecystectomy on a porcine organ model. Two expert surgeons rated the endoscopic videos of the cholecystectomies blinded to training status of the participants.

The selection of participants might influence the results of a transfer of skills study. We chose to use medical students for mainly two reasons. First to have a homogenous group of participants, and second that it was easier to recruit a large enough number of participants among last year medical students. The disadvantage might be that medical students might differ more in innate motor skills abilities than surgical residents who have chosen to become surgeons.

Our clinical setting was a porcine organ model, and not a real clinical setting. This might have influenced the results.

The expert surgeons used the Global Operative Assessment of Laparoscopic Skills (GOALS) tool (Vassiliou, 2005) to rate the videos. The rating scale consist of five components: *depth perception*, *bimanual dexterity*, *efficiency*, *tissue handling* and *autonomy*. Vassilou et al. (2005) performed a validation study on the GOALS scale indicating that it was reliable and valid under their setting: an operating room. We used the GOALS scale to rate endoscopic videos leaving out the *autonomy* component, assuming that the other four components would be valid also in our setting. One might argue that it would have strengthen our study if we would have performed a validation study, investigating that assumption, prior to it. Brinkman et al. made the same assumption as we did and used the GOALS scale in a similar study to rate videos (Brinkmann, 2017).

Appropriate statistical tests were found for each study using primarily nonparametric tests.

The role of simulation-based training in surgical education

Simulation-based training has come to stay in surgical education. The advantages are numerous and are often presented together with challenges within surgical education that it can supposedly solve (Fry, 2011; Prentice, 2012; Stefanidis, 2019). Three of the most common challenges, where simulation-based training and assessment have been proposed as solutions, are:

1) Reliance on patients for training. Simulation-based training is regarded as a risk-free training alternative outside of the operating room
2) A decrease in surgical related exposure due to reduced working hours and other organizational changes. Simulation-based surgical exposure might replace reduced real-world exposure.
3) Lack of objective quality assurance and assessment of surgical skills. Competency-based training and assessment using simulators to assess surgical skills objectively might assure qualified surgeons having the needed qualifications.

Risk-free training outside of the operating room

Moving training out of the operating room and into the skills lab is probably the most important advantage of simulation-based training. The increase in incidents and accidents with the introduction of laparoscopy and the Bristol case are examples of difficult ethical aspects of surgical education. It is not acceptable to put patients at risk while surgeons train. If surgeons can train outside of the operating room using simulators, thereby not putting patients at risk, that ought to be a better alternative. There are, however, some ifs and buts that are often not mentioned nor discussed. First of all, it is important to remember that, in a well-functioning apprenticeship model the apprentice is allowed to perform tasks under close supervision with increasingly difficulty as he or she proves fit for them (Polavarapu, 2013; Prentice, 2012). Thus, training in the OR doesn't necessarily put the patients at a higher risk. But the apprenticeship model is under pressure with demands on efficiency, surgical procedures with greater complexity, and organisational changes (Fry, 2011; Prentice, 2012). Therefore, moving training out of the busy OR is a good idea, given that simulation-based training can replace training in the OR.

There are two ways of looking at whether simulation-based training can replace training in the OR, and if yes, to what extent. One way is to perform studies investigating transfer of skills: do skills acquired (and measured) on the simulator transfer to the OR? The other is to study and reflect upon the skills that are learnt in the OR and the skills that are learnt on a simulator, i.e. what are the differences between training in the OR and in the skills lab?

To understand how simulation-based training can help improve surgical education, it is important to reflect upon the differences between training in the OR and the skills lab: What is learned in the OR and what is learned in the skills lab? A surgical trainee might learn to suture laparoscopically on a simulator, but that doesn't mean that he is ready to perform a full procedure as a leading operator. This was briefly reflected upon in the paper on procedural VR simulation (Våpenstad, 2013a). Being a surgeon in the OR is far more complex than performing a surgical task on a simulator, and it should be emphasized that simulation-based training is an *adjunct* to apprenticeship-based surgical education. Many papers add the word *"adjunct to"* when referring to simulation-based training, others present challenges with today surgical education and introduce simulation-based training as *the* solution.

Despite the fact that surgery is taught through an apprenticeship model, the typical tradition has been to see knowledge and skills as an individual enterprise where the individual is the carrier and receiver of knowledge (Bleakley, 2006), and where learning is seen as an activity apart. Lave and Wenger questions these assumptions, and argues, that learning is situated and *"an integral and inseparable aspect of social practice"* (Lave, 1991b, p. 31). The intention of surgical education is to transfer skills and knowledge alongside with values and principals so that the trainee can work effectively in the surgical environment. The learning theory of 'situated learning', as a result of Lave and Wenger's investigation into apprenticeships, takes into account the relational and social aspects of learning (Lave, 1991b). They view learning as *'legitimate peripheral participation'* where actors negotiate their participation through social involvement and thereby learn how to become and act as members of the *'community of practice'* (Lave, 1991b; Wenger, 1999). As the surgical trainees demonstrate step-wise increased proficiency, they are allowed to do more difficult tasks and, in the end, perform full procedures under supervision. The theory of situated learning emphasises merited involvement in communities of practice, where the trainees decode either consciously or non-consciously tacit clues, and where simulators might help the trainees acquire basic skills and (personal) knowledge that will more easily grant them a place in the community of practice. The surgical community will also spend less time guiding and correcting the trainees if they have acquired basic skills and knowledge on simulators before training in the operating room.

It is important to remember that simulation-based training is not proven per se to have a positive effect on surgical skills. Just training on a simulator will not necessarily increase the candidate's surgical skills. There are studies that show transfer of skills to the operating room (Ahlberg, 2007; Nagendran, 2013; Seymour, 2008), but these are valid for that given simulator set-up. There are also studies showing the opposite, like our study of transfer of skills from the LapSim® VR simulator with haptic handles, which indicated that training on that simulator had a negative transfer of skills (Våpenstad, 2017). Validity evidence should be sought prior to use, and more evidence is needed for uses with higher-stakes (Korndorffer, 2010).

When looking at simulators it is important to keep in mind the difference between physical resemblance (fidelity) and functional task alignment (Hamstra, 2014; Wanzel, 2002). Physical resemblance is whether the simulator looks and feels like a surgical environment, and functional task alignment is whether training on it actually transfer to the operating room. It is functional task alignment that matters, but this is influenced by the simulator's physical resemblance. A simulator has to reproduce essential parts of the surgical environment, and if it does, that will be exposed through scientific studies investigating functional task alignment. To see what are essential parts of the surgical environment is not always easy, and what if a simulator reproduces the surgical scene with great fidelity but lack physical resemblance on other aspects? Haptic feedback is a good example. It is an essential part of laparoscopy, i.e. high physical resemblance is essential, but it has been difficult to achieve haptic feedback with high physical resemblance for VR simulators (Lamata, 2006a; Lamata, 2006b; van der Meijden, 2009; Våpenstad, 2013b; Våpenstad, 2017; Våpenstad, 2013c). Thus, proof of functional task alignment can be difficult to find (Våpenstad, 2017), regardless of the fidelity of the virtual surgical scene, if the haptic feedback's physical resemblance is too low.

Simulators and especially VR simulators have seen large technological advances the last decades and will probably continue to do so. The challenge for engineers, will be, together with clinical experts to translate real surgical environments into bits and bytes (Gorman, 2000; Prentice, 2012): technological solutions that will have to be integrated and validated in clinical settings. Further, technology is important but not the only aspect of successful training on simulators: e.g. it matters how the training set-up is organized, with or without expert surgeons available for guidance, or simply motivation and training strategies with the candidate (Våpenstad, 2013a).

Currently, simulators are not capable of reconstructing all aspects of performing surgery in a surgical community, and these aspects, whether they are called situated learning (Dimitriadis, 2014; Lave, 1991a; Lave, 1991b; Sadideen, 2012b; Wenger, 1999), non-formal or tacit knowledge (Engel, 2008; Eraut, 2000; Fry, 2011; Prentice, 2012), or normalizing technologies of self (Jaye, 2006; Jaye, 2010), should to be kept in mind when looking at surgical education as a whole. Simulation-based training therefore ought to be presented as an adjunct to apprenticeship-based training, and consequences regarding e.g. transfer of values and principals should be considered when moving part of surgical education out of the operating room into skills labs.

Reduced surgical exposure

Simulation-based training is often proposed as a solution to the decreased operative volume and reduced surgical exposure due to reduced working-hours and other organisational changes within surgical wards. The reason for reduced working hours and the following challenges (and benefits) have created much debate (Glomsaker, 2009; Kahol, 2008; Leff, 2007; Shanafelt, 2010; Underwood, 2003). The debate has opponents on either side, and the point here is not whether working-hours should be reduced or what the potential challenges (or benefits) are, but whether simulation-based training is an efficient tool to compensate for reduced working-hours and a subsequent reduction in surgical exposure. One might argue that if surgical trainees works less hours, then they also have less time to train on simulators, and moving them out of the OR and into the skills lab to train, moves them even more away from surgical exposure. But simulation-based training on psychomotor skills might be more efficient than the equivalent in the OR, as a task can be repeated numerous times with limited time spent on preparations. The cost of operating time is high and if trainees can prepare on simulators before procedures, thereby reducing operating time that would reduce overall costs (Louridas, 2015). If simulators are available at the ward, training on them can be done in between and during quiet moments when being on duty. The skills lab environment is also a more controlled environment with the possibility of formative feedback facilitating e.g. training following the principles of deliberate practice (Crochet, 2011). VR simulators simulating procedures can also be used to train on procedures that do not occur regularly, either due to the procedure itself being rare or the hospital being small (Bell, 2009). On the other hand, training on a simulator is generally seen as a low stake, whereas performing the same task during a procedure in the operating room is a high stake. Prentice evokes the differences between low stakes and high stakes, and that when experiencing a high-stake situation, human beings are more likely to remember and thus acquire the lesson more efficiently (Prentice, 2012).

In conclusion, simulation-based training can be more flexible than acquiring skills in the OR. However, it is still time consuming, and as mentioned previously simulation-based training is not equal to surgical related exposure in the OR. It should be regarded as an adjunct to the apprentice-ship based model, and can thus only partly solve challenges related to reduced surgical exposure.

Quality assurance through simulation-based assessment

Another advantage of simulators is a quality assurance of surgeons' qualifications through competency-based education using simulators to assess skills objectively (Stefanidis, 2019). The increase in incidents and accidents with the introduction of laparoscopy and landmark cases like the Bristol case, made healthcare management and other stakeholders eager to quality assure that surgeons acquire the necessary skills and competencies to perform surgery safely through objective assessment (Fry, 2011). Many simulators have assessment tools that can measure skills, and that can be used to assure that surgeons have specific skills. Whether a simulator can measure *surgical* skills or not is investigated in what was called a construct validity test in the old paradigm or relationship to other variables in Messick's framework (Borgersen, 2018; Gallagher, 2003; Sweet, 2010), and is what we did in paper 3 and 5 (Våpenstad, 2013b; Våpenstad, 2019).

If a simulator can be used to measure surgical skills it can be used to give a quality mark through a certification or help in the selection process of surgical trainees (Gardner, 2016; Louridas, 2015). In a selection process, the simulator have to measure innate abilities whereas in a quality assurance process it has to measure actual surgical skills, in both cases parameters and cut-off values must be investigated (Goldenberg, 2017). In three studies investigating different motion parameters and how they can be used to differentiate between expert surgeons and novices, we found that e.g. expert surgeons had a greater bimanual dexterity than less experienced surgeons (Hofstad, 2017; Hofstad, 2013; Våpenstad, 2019). Work is still in progress on which parameters are good indicators on innate abilities or surgical skills (Thinggaard, 2016), and to collect validity evidence on simulators that can be used to measure them (Louridas, 2015; Stefanidis, 2019).

The Fundamentals of Laparoscopic Surgery® technical skills tests are probably the most used and the tests that have the largest volume of evidence to support their use (Bilgic, 2018; Okrainec, 2011; Peters, 2004; Stefanidis, 2019). The Fundamentals of Laparoscopic Surgery® includes both technical and cognitive skills and addresses thus several aspects of competencies that are needed as a surgeon. There are several ways of assessing skills and competencies in addition to simulation-based assessment such as The Global Operative Assessment of Laparoscopic Skills (GOALS) tool (Vassiliou, 2005) and the Objective Structured Assessment of Technical Skills (OSATS) scale (Martin, 1997; Niitsu, 2013). These scales are usually developed for assessment in an operative setting, and can be seen as complementary to simulation-based assessment in the skills lab. A combination of tests that can evaluate different kinds of surgical competencies are probably the best (Vassiliou, 2011). It should although be kept in mind, that verifying a trainees acquisition of specific technical skills, is far easier than verifying e.g. their acquisition of in-depth surgical anatomical perception, decision making competencies and respect for patient bodies (Prentice, 2012).

Possible advantages and drawbacks of simulation-based training are presented in Figure 6.

Advantages:

- Training without reliance on patients
- Objective Skills Assessment of errors, motions, forces
- Formative and Summative feedback
- Flexible, limited need for preparation
- Removes pressure from the busy OR
- Repetitive training on rare procedures

Drawbacks:

- Low stake
- A simplification of the real world
- Requires evidence of validity before use
- Aspects such as haptic feedback technically challenging to simulate
- Equipment can be expensive

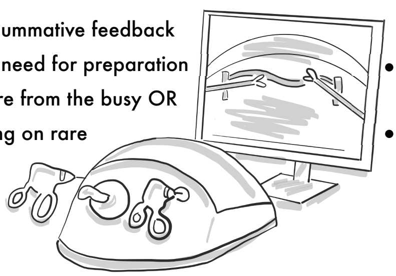

Figure 6: Possible advantages and drawbacks with simulation-based training and assessment, courtesy: Cecilie Våpenstad

Scientific studies are generally founded on questions of interests or objectives, most often high-level questions such as "do training on simulators transfer to a clinical setting?". High-level questions are usually those that scientists are most eager to answer. Unfortunately, feasibility and practical aspects often constrain study design. High-level questions are, thus, often answered through a generalization of a feasible study investigating a lower-level question, such as "do surgeons prefer this simulator with haptic feedback over this simulator without it?".

Transfer of skills from simulator training to clinical performance, the 1st objective, was investigated in paper 4. We did not find transfer of skills, probably due to unrealistic simulation of the haptic feedback. Although, our study is valid for the specific set-up, the finding that an unrealistic simulator has negative training effect is important in general. The negative training effect incites carefulness and that validity should be investigated prior to use.

in paper 3 and 5 we investigated whether documented simulator competence correspond with clinical competence, the 2nd objective, for two different simulators: a VR simulator and a box trainer. We partly found correspondence for the two, and although our studies are valid for the specific set-ups some aspects can be generalized, such as the importance of describing the set-up including the handles, and the timing of an assessment task.

The characteristics of simulation-based training, the 3rd objective, was explored in all of the papers, and particularly in paper 1 and 2. We proposed a taxonomy for VR simulators dividing into fidelity resources, teaching resources and management resources. Haptic feedback was found to be important, but the handles that we investigated did not simulate the sensations of haptic feedback well enough.

The backdrop for this thesis was the National Advisory Board for Advanced Laparoscopy's (NSALK) effort to implement VR simulators and box trainers into surgical education in Norway. NSALK actively uses their simulators at their courses. Some VR simulators have been replaced by simulators with different handles. These simulators simulate haptics pushing and pulling the instruments instead of applying a force on them. We used the box trainer to assess surgical trainees over a time period of three years in Mid-Norway, but were unable to continue due to financial and practical constraints. There is an ongoing effort at the center, and we e.g. also tested low-cost simulators that we lend out to trainees and tested them again after two months (not presented in this thesis). The study was ended due to reliability issues with the simulator. Practical, political and financial aspects regulate the implementation of simulation-based training and assessment. The main focus of this thesis was to explore and validate simulation-based training and assessment with both practical intentions for NSALK and scientific intentions for the wider surgical education community.

Main conclusions

Simulation-based training and assessment are important adjuncts to the traditional apprenticeship model, given that the simulators exhibit enough validity evidence for the intended use. Through the five papers that this PhD is based on, I have, together with my co-authors, explored and validated simulation-based training and assessment. Through the literature review and questionnaire sent out to simulator manufacturers we found that several simulated procedures are available for training, but only a limited number of them have been part of a validation study. We further proposed a taxonomy of VR simulators dividing into fidelity resources, teaching resources and management resources. We did not find transfer of skills, and we found that the simulators tested were capable of distinguishing between levels of surgical skills for some tasks and some parameters, whereas others failed to do so. We further found that assessment tasks should be adapted to level of training. Presented in three of the papers, we found that haptic feedback is an important part of VR simulation, and an aspect that it has been difficult to simulate sufficiently well so far. Although an important part of VR simulation, we found that the handles used were often poorly described in the literature.

Although simulation-based training and assessment is slowly increasing in use, there is still a need for further exploration, development and validation. And, last but not least, the deployment of simulation-based training is dependent on both organizational and political aspects, that provide room for a good balance between simulation-based and apprenticeship-based surgical education.

References:

ACGME. (2019). *ACGME Core Competencies*. Retrieved 29.05.2019, from http://health.usf.edu/medicine/surgery/general/residency/acgme

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological bulletin, 101*(1), 41. https://doi.org/10.1037/0033-2909.101.1.41

Aggarwal, R., Grantcharov, T., Moorthy, K., Hance, J., & Darzi, A. (2006). A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *Am J Surg, 191*(1), 128-133. https://doi.org/10.1016/j.amjsurg.2005.10.014

Ahlberg, G., Enochsson, L., Gallagher, A. G., Hedman, L., Hogman, C., McClusky, D. A., 3rd, et al. (2007). Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg, 193*(6), 797-804. https://doi.org/10.1016/j.amjsurg.2006.06.050

APA, A. N. (2014). The Standards for Educational and Psychological Testing. Washington, USA: Joint Committee on Standards for Educational and Psychological Testing: American Psychological Association, American Educational Research Association and National Council on Measurement in Education.

Aucar, J. A., Groch, N. R., Troxel, S. A., & Eubanks, S. W. (2005). A review of surgical simulation with attention to validation methodology. *Surg Laparosc Endosc Percutan Tech, 15*(2), 82-89. https://doi.org/10.1097/01.sle.0000160289.01159.0e

Bann, S. D., Khan, M. S., & Darzi, A. W. (2003). Measurement of surgical dexterity using motion analysis of simple bench tasks. *World J Surg, 27*(4), 390-394. https://doi.org/10.1007/s00268-002-6769-7

Basdogan, C., De, S., Kim, J., Muniyandi, M., Kim, H., & Srinivasan, M. A. (2004). Haptics in minimally invasive surgical simulation and training. *IEEE Comput Graph Appl, 24*(2), 56-64. https://doi.org/10.1109/MCG.2004.1274062

Behrendt, C. A., Sedrakyan, A., Riess, H. C., Heidemann, F., Kolbel, T., Petersen, J., et al. (2017). Short-term and long-term results of endovascular and open repair of abdominal aortic aneurysms in Germany. *J Vasc Surg, 66*(6), 1704-1711 e1703. https://doi.org/10.1016/j.jvs.2017.04.040

Bell, R. H., Jr., Biester, T. W., Tabuenca, A., Rhodes, R. S., Cofer, J. B., Britt, L. D., et al. (2009). Operative experience of residents in US general surgery programs: a gap between expectation and experience. *Ann Surg, 249*(5), 719-724. https://doi.org/10.1097/SLA.0b013e3181a38e59

Berguer, R., Smith, W. D., & Chung, Y. H. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surg Endosc, 15*(10), 1204-1207. https://doi.org/10.1007/s004640080030

Bholat, O. S., Haluck, R. S., Murray, W. B., Gorman, P. J., & Krummel, T. M. (1999). Tactile feedback is present during minimally invasive surgery. *Journal of the American College of Surgeons, 189*(4), 349-355. https://doi.org/10.1016/S1072-7515(99)00184-2

Bilgic, E., Kaneva, P., Okrainec, A., Ritter, E. M., Schwaitzberg, S. D., & Vassiliou, M. C. (2018). Trends in the Fundamentals of Laparoscopic Surgery(R) (FLS) certification exam over the past 9 years. *Surg Endosc, 32*(4), 2101-2105. https://doi.org/10.1007/s00464-017-5907-0

Birkmeyer, J. D., Finks, J. F., O'Reilly, A., Oerline, M., Carlin, A. M., Nunn, A. R., et al. (2013). Surgical skill and complication rates after bariatric surgery. *N Engl J Med, 369*(15), 1434-1442. https://doi.org/10.1056/NEJMsa1300625

Bleakley, A. (2006). Broadening conceptions of learning in medical education: the message from teamworking. *Medical Education, 40*(2), 150-157. https://doi.org/10.1111/j.1365-2929.2005.02371.x

Borgersen, N. J., Naur, T. M. H., Sorensen, S. M. D., Bjerrum, F., Konge, L., Subhi, Y., et al. (2018). Gathering Validity Evidence for Surgical Simulation: A Systematic Review. *Ann Surg, 267*(6), 1063-1068. https://doi.org/10.1097/SLA.0000000000002652

Botden, S. M., Buzink, S. N., Schijven, M. P., & Jakimowicz, J. J. (2007). Augmented versus virtual reality laparoscopic simulation: what is the difference? A comparison of the ProMIS augmented reality laparoscopic simulator versus LapSim virtual reality laparoscopic simulator. *World J Surg, 31*(4), 764-772. https://doi.org/10.1007/s00268-006-0724-y

Botden, S. M., Torab, F., Buzink, S. N., & Jakimowicz, J. J. (2008). The importance of haptic feedback in laparoscopic suturing training and the additive value of virtual reality simulation. *Surg Endosc, 22*(5), 1214-1222. https://doi.org/10.1007/s00464-007-9589-x

Brinkmann, C., Fritz, M., Pankratius, U., Bahde, R., Neumann, P., Schlueter, S., et al. (2017). Box- or Virtual-Reality Trainer: Which Tool Results in Better Transfer of Laparoscopic Basic Skills?-A Prospective Randomized Trial. *J Surg Educ, 74*(4), 724-735. https://doi.org/10.1016/j.jsurg.2016.12.009

Buzink, S. N., Goossens, R. H., De Ridder, H., & Jakimowicz, J. J. (2010). Training of basic laparoscopy skills on SimSurgery SEP. *Minim Invasive Ther Allied Technol, 19*(1), 35-41. https://doi.org/10.3109/13645700903384468

Cao, C. G., Zhou, M., Jones, D. B., & Schwaitzberg, S. D. (2007). Can surgeons think and operate with haptics at the same time? *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract, 11*(11), 1564-1569. https://doi.org/10.1007/s11605-007-0279-8

Carter, F. J., Schijven, M. P., Aggarwal, R., Grantcharov, T., Francis, N. K., Hanna, G. B., et al. (2005). Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc, 19*(12), 1523-1532. https://doi.org/10.1007/s00464-005-0384-2

Cates, C. U., Patel, A. D., & Nicholson, W. J. (2007). Use of virtual reality simulation for mission rehearsal for carotid stenting. *JAMA, 297*(3), 265-266. https://doi.org/10.1001/jama.297.3.265-b

Chmarra, M. K., Dankelman, J., van den Dobbelsteen, J. J., & Jansen, F. W. (2008). Force feedback and basic laparoscopic skills. *Surg Endosc, 22*(10), 2140-2148. https://doi.org/10.1007/s00464-008-9937-5

Coles, T. R., Meglan, D., & John, N. W. (2011). The Role of Haptics in Medical Training Simulators: A Survey of the State of the Art. *Ieee Transactions on Haptics, 4*(1), 51-66. https://doi.org/10.1109/Toh.2010.19

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med, 119*(2), 166 e167-116. https://doi.org/10.1016/j.amjmed.2005.10.036

Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J., & Hatala, R. (2013). Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med, 88*(6), 872-883. https://doi.org/10.1097/ACM.0b013e31828ffdcf

Cook, D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract, 19*(2), 233-250. https://doi.org/10.1007/s10459-013-9458-4

Crochet, P., Aggarwal, R., Dubb, S. S., Ziprin, P., Rajaretnam, N., Grantcharov, T., et al. (2011). Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. *Ann Surg, 253*(6), 1216-1222. https://doi.org/10.1097/SLA.0b013e3182197016

Cuschieri, A. (1995). Whither minimal access surgery: tribulations and expectations. *Am J Surg, 169*(1), 9-19. https://doi.org/10.1016/S0002-9610(99)80104-4

Danila, R., Gerdes, B., Ulrike, H., Dominguez Fernandez, E., & Hassan, I. (2009). Objective evaluation of minimally invasive surgical skills for transplantation. Surgeons using a virtual reality simulator. *Chirurgia (Bucur), 104*(2), 181-185. https://doi.org/10.1016/j.ijsu.2014.11.014

Davis, G. R., Illig, K. A., Yang, G., Nguyen, T. H., & Shames, M. L. (2014). An approach to EVAR simulation using patient specific modeling. *Ann Vasc Surg, 28*(7), 1769-1774. https://doi.org/10.1016/j.avsg.2014.05.007

Desender, L., Rancic, Z., Aggarwal, R., Duchateau, J., Glenck, M., Lachat, M., et al. (2013). Patient-specific rehearsal prior to EVAR: a pilot study. *Eur J Vasc Endovasc Surg, 45*(6), 639-647. https://doi.org/10.1016/j.ejvs.2013.03.006

Desender, L., Van Herzeele, I., Lachat, M., Duchateau, J., Bicknell, C., Teijink, J., et al. (2017a). A Multicentre Trial of Patient specific Rehearsal Prior to EVAR: Impact on Procedural Planning and Team Performance. *Eur J Vasc Endovasc Surg, 53*(3), 354-361. https://doi.org/10.1016/j.ejvs.2016.12.018

Desender, L. M., Van Herzeele, I., Rancic, Z., Bicknell, C., Zairis, I., Vermassen, F. E., et al. (2017b). Patient-Specific Simulation of Endovascular Thoracic Aortic Repair: Initial Experience. *Ann Thorac Surg, 104*(1), 336-341. https://doi.org/10.1016/j.athoracsur.2016.09.066

Dimitriadis, P., Iyer, S., & Evgeniou, E. (2014). Learning in the Surgical Community of Practice. *Medical Science Educator, 24*(2), 211-214. https://doi.org/10.1007/s40670-014-0042-1

Downing, S. M., & Haladyna, T. M. (2004). Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ, 38*(3), 327-333. https://doi.org/10.1046/j.1365-2923.2004.01777.x

Duffy, A. J., Hogle, N. J., McCarthy, H., Lew, J. I., Egan, A., Christos, P., et al. (2005). Construct validity for the LAPSIM laparoscopic surgical simulator. *Surg Endosc, 19*(3), 401-405. https://doi.org/10.1007/s00464-004-8202-9

Dyer, C. (1999). Bristol case surgeon claimed to have been on "learning curve. *British Medical Journal, 319*(7223), 1456. https://doi.org/10.1136/bmj.319.7223.1456f

El Saddik, A. (Ed.). (2012). *Haptics Rendering and Applications*: InTech.

Engel, P. J. H. (2008). Tacit knowledge and visual expertise in medical diagnostic reasoning: Implications for medical education. *Medical Teacher, 30*(7), E184-E188. https://doi.org/10.1080/01421590802144260

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology, 70*(1), 113-136. https://doi.org/10.1348/000709900158001

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med, 79*(10 Suppl), S70-81. https://doi.org/10.1111/j.1553-2712.2008.00227.x

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med, 15*(11), 988-994. https://doi.org/10.1111/j.1553-2712.2008.00227.x

Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: A perspective from the expert-performance approach with deliberate practice. *Academic Medicine, 90*(11), 1471-1486. https://doi.org/10.1097/ACM.0000000000000939

Fairhurst, K., Strickland, A., & Maddern, G. (2011). The LapSim virtual reality simulator: promising but not yet proven. *Surg Endosc, 25*(2), 343-355. https://doi.org/10.1007/s00464-010-1181-0

Fitts, P., & Posner, M. (1967). *Human Performance*. Oxford, England: Brooks/Cole.

Fry, H., & Kneebone, R. (2011). *Surgical Education: Theorising an emerging domain* (Vol. 2). Dordrecht, The Netherlands: Springer Science & Business Media.

Fuchs, K. H. (2002). Minimally invasive surgery. *Endoscopy, 34*(2), 154-159. https://doi.org/10.1055/s-2002-19857

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., et al. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg, 241*(2), 364-372. https://doi.org/10.1097/01.sla.0000151982.85062.80

Gallagher, A. G., Ritter, E. M., & Satava, R. M. (2003). Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc, 17*(10), 1525-1529. https://doi.org/10.1007/s00464-003-0035-4

Ganpule, A., Chhabra, J. S., & Desai, M. (2015). Chicken and porcine models for training in laparoscopy and robotics. *Curr Opin Urol, 25*(2), 158-162. https://doi.org/10.1097/MOU.0000000000000139

Gardner, A. K., Ritter, E. M., Paige, J. T., Ahmed, R. A., Fernandez, G., & Dunkin, B. J. (2016). Simulation-Based Selection of Surgical Trainees: Considerations, Challenges, and Opportunities. *J Am Coll Surg, 223*(3), 530-536. https://doi.org/10.1016/j.jamcollsurg.2016.05.021

Glomsaker, T. B., & Soreide, K. (2009). Surgical training and working time restriction. *Br J Surg, 96*(4), 329-330. https://doi.org/10.1002/bjs.6595

Goldenberg, M., & Lee, J. Y. (2018). Surgical Education, Simulation, and Simulators-Updating the Concept of Validity. *Curr Urol Rep, 19*(7), 52. https://doi.org/10.1007/s11934-018-0799-7

Goldenberg, M. G., Garbens, A., Szasz, P., Hauer, T., & Grantcharov, T. P. (2017). Systematic review to establish absolute standards for technical performance in surgery. *Br J Surg, 104*(1), 13-21. https://doi.org/10.1002/bjs.10313

Goodwin, C. (1994). Professional Vision. *American Anthropologist, 96*(3), 606-633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Gorman, P., Meier, A. H., Rawn, C., & Krummel, T. M. (2000). The future of medical education is no longer blood and guts, it is bits and bytes,. *The American Journal of Surgery, 180*(5), 353-356. https://doi.org/10.1016/S0002-9610(00)00514-6

Grimm, J. (2005). Interactive real-time simulation of an endoscopic polyp removal. In J. Westwood, R. Haluck & H. Hoffman (Eds.), *Medicine Meets Virtual Reality 13: The Magical Next Becomes the Medical Now* (pp. 160-163). IOS Press.

Gruppen, L. D., Mangrulkar, R. S., & Kolars, J. C. (2012). The promise of competency-based education in the health professions for improving global health. *Hum Resour Health, 10*, 43. https://doi.org/10.1186/1478-4491-10-43

Guedes, H. G., Camara Costa Ferreira, Z. M., Ribeiro de Sousa Leao, L., Souza Montero, E. F., Otoch, J. P., & Artifon, E. L. A. (2019). Virtual reality simulator versus box-trainer to teach minimally invasive procedures: A meta-analysis. *Int J Surg, 61*, 60-68. https://doi.org/10.1016/j.ijsu.2018.12.001

Gurusamy, K. S., Aggarwal, R., Palanivelu, L., & Davidson, B. R. (2009). Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev, 2009*(1), CD006575. https://doi.org/10.1002/14651858.CD006575.pub2

Gurusamy, K. S., Nagendran, M., Toon, C. D., & Davidson, B. R. (2014). Laparoscopic surgical box model training for surgical trainees with limited prior laparoscopic experience. *Cochrane Database Syst Rev, 3*, CD010478. https://doi.org/10.1002/14651858.CD010478.pub2

Hagelsteen, K., Johansson, R., Ekelund, M., Bergenfelz, A., & Anderberg, M. (2019). Performance and perception of haptic feedback in a laparoscopic 3D virtual reality simulator. *Minim Invasive Ther Allied Technol*, 1-8. https://doi.org/10.1080/13645706.2018.1539012

Hagelsteen, K., Sevonius, D., Bergenfelz, A., & Ekelund, M. (2016). Simball Box for Laparoscopic Training With Advanced 4D Motion Analysis of Skills. *Surg Innov, 23*(3), 309-316. https://doi.org/10.1177/1553350616628678

Hamdorf, J. M., & Hall, J. C. (2000). Acquiring surgical skills. *Br J Surg, 87*(1), 28-37. https://doi.org/10.1046/j.1365-2168.2000.01327.x

Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Acad Med, 89*(3), 387-392. https://doi.org/10.1097/ACM.0000000000000130

Hennessey, I. A., & Hewett, P. (2013). Construct, concurrent, and content validity of the eoSim laparoscopic simulator. *J Laparoendosc Adv Surg Tech A, 23*(10), 855-860. https://doi.org/10.1089/lap.2013.0229

Himal, H. S. (2002). Minimally invasive (laparoscopic) surgery. *Surg Endosc, 16*(12), 1647-1652. https://doi.org/10.1007/s00464-001-8275-7

Hislop, S. J., Hedrick, J. H., Singh, M. J., Rhodes, J. M., Gillespie, D. L., Johansson, M., et al. (2009). Simulation case rehearsals for carotid artery stenting. *Eur J Vasc Endovasc Surg, 38*(6), 750-754. https://doi.org/10.1016/j.ejvs.2009.08.011

Hofstad, E. F., Vapenstad, C., Bo, L. E., Lango, T., Kuhry, E., & Marvik, R. (2017). Psychomotor skills assessment by motion analysis in minimally invasive surgery on an animal organ. *Minim Invasive Ther Allied Technol, 26*(4), 240-248. https://doi.org/10.1080/13645706.2017.1284131

Hofstad, E. F., Vapenstad, C., Chmarra, M. K., Lango, T., Kuhry, E., & Marvik, R. (2013). A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. *Surg Endosc, 27*(3), 854-863. https://doi.org/10.1007/s00464-012-2524-9

Horeman, T., Dankelman, J., Jansen, F. W., & van den Dobbelsteen, J. J. (2014). Assessment of laparoscopic skills based on force and motion parameters. *IEEE Trans Biomed Eng, 61*(3), 805-813. https://doi.org/10.1109/TBME.2013.2290052

Horeman, T., Meijer, E. J., Harlaar, J. J., Lange, J. F., van den Dobbelsteen, J. J., & Dankelman, J. (2013). Force sensing in surgical sutures. *PLoS One, 8*(12), e84466. https://doi.org/10.1371/journal.pone.0084466

Horeman, T., Rodrigues, S. P., Jansen, F. W., Dankelman, J., & van den Dobbelsteen, J. J. (2010). Force measurement platform for training and assessment of laparoscopic skills. *Surg Endosc, 24*(12), 3102-3108. https://doi.org/10.1007/s00464-010-1096-9

Horeman, T., Rodrigues, S. P., Jansen, F. W., Dankelman, J., & van den Dobbelsteen, J. J. (2012). Force Parameters for Skills Assessment in Laparoscopy. *IEEE Trans Haptics, 5*(4), 312-322. https://doi.org/10.1109/TOH.2011.60

Huber, T., Paschold, M., Hansen, C., Wunderling, T., Lang, H., & Kneist, W. (2017). New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff. *Surg Endosc, 31*(11), 4472-4477. https://doi.org/10.1007/s00464-017-5500-6

Huber, T., Wunderling, T., Paschold, M., Lang, H., Kneist, W., & Hansen, C. (2018). Highly immersive virtual reality laparoscopy simulation: development and future aspects. *Int J Comput Assist Radiol Surg, 13*(2), 281-290. https://doi.org/10.1007/s11548-017-1686-2

Jaye, Egan, T., & Parker, S. (2006). 'Do as I say, not as I do': Medical Education and Foucault's Normalizing Technologies of Self. *Anthropology & Medicine, 13*(2), 141-155. https://doi.org/10.1080/13648470600738450

Jaye, C., Egan, T., & Smith-Han, K. (2010). Communities of clinical practice and normalising technologies of self: learning to fit in on the surgical ward. *Anthropology & medicine, 17*(1), 59-73. https://doi.org/10.1080/13648470903569388

Kahol, K., Leyba, M. J., Deka, M., Deka, V., Mayes, S., Smith, M., et al. (2008). Effect of fatigue on psychomotor and cognitive skills. *Am J Surg, 195*(2), 195-204. https://doi.org/10.1016/j.amjsurg.2007.10.004

Katz, P. (1999). *The scalpel's edge: The culture of surgeons*: Allyn and Bacon Needham Heights.

Kelley, W. E., Jr. (2008). The evolution of laparoscopy and the revolution in surgery in the decade of the 1990s. *JSLS, 12*(4), 351-357.

Kim, H. K., Rattner, D. W., & Srinivasan, M. A. (2004). Virtual-reality-based laparoscopic surgical training: the role of simulation fidelity in haptic feedback. *Comput Aided Surg, 9*(5), 227-234. https://doi.org/10.3109/10929080500066997

Korndorffer, J. R., Jr., Kasten, S. J., & Downing, S. M. (2010). A call for the utilization of consensus standards in the surgical education literature. *Am J Surg, 199*(1), 99-104. https://doi.org/10.1016/j.amjsurg.2009.08.018

Lahanas, V., Georgiou, E., & Loukas, C. (2016). Surgical Simulation Training Systems: Box Trainers, Virtual Reality and Augmented Reality Simulators. *Int J Adv Robot Automn, 1*(2), 1-9. https://doi.org/10.15226/2473-3032/1/2/00109

Lamata, P., Gomez, E. J., Bello, F., Kneebone, R. L., Aggarwal, R., & Lamata, F. (2006a). Conceptual framework for laparoscopic VR simulators. *Ieee Computer Graphics and Applications, 26*(6), 69-79. https://doi.org/10.1109/MCG.2006.125

Lamata, P., Gomez, E. J., Sanchez-Margallo, F. M., Lamata, F., Antolin, M., Rodriguez, S., et al. (2006b). Study of laparoscopic forces perception for defining simulation fidelity. *Stud Health Technol Inform, 119*, 288-292.

Lave, J. (1991a). Situating learning in communities of practice. *Perspectives on socially shared cognition, 2*, 63-82.

Lave, J., & Wenger, E. (1991b). *Situated learning: Legitimate peripheral participation*: Cambridge university press.

Leff, D., Aziz, O., & Darzi, A. (2007). Trucks, planes, and scalpels: is there an evidence-based approach to surgeons' working hours? *Arch Surg, 142*(9), 817-820. https://doi.org/10.1001/archsurg.142.9.817

Li, M. M., & George, J. (2017). A systematic review of low-cost laparoscopic simulators. *Surg Endosc, 31*(1), 38-48. https://doi.org/10.1007/s00464-016-4953-3

Litynski, G. S. (1997). Laparoscopy--the early attempts: spotlighting Georg Kelling and Hans Christian Jacobaeus. *JSLS, 1*(1), 83-85.

Liu, A., Tendick, F., Cleary, K., & Kaufmann, C. (2003). A survey of surgical simulation: Applications, technology, and education. *Presence-Teleoperators and Virtual Environments, 12*(6), 599-614. https://doi.org/10.1162/105474603322955905

Loukas, C., Lahanas, V., & Georgiou, E. (2013). An integrated approach to endoscopic instrument tracking for augmented reality applications in surgical simulation training. *Int J Med Robot, 9*(4), e34-51. https://doi.org/10.1002/rcs.1485

Louridas, M., Szasz, P., de Montbrun, S., Harris, K. A., & Grantcharov, T. P. (2015). Can We Predict Technical Aptitude? A Systematic Review. *Ann Surg*. https://doi.org/10.1097/SLA.0000000000001283

Malik, H. H., Darwood, A. R., Shaunak, S., Kulatilake, P., El-Hilly, A. A., Mulki, O., et al. (2015). Three-dimensional printing in surgery: a review of current surgical applications. *J Surg Res, 199*(2), 512-522. https://doi.org/10.1016/j.jss.2015.06.051

Martin, J. A., Regehr, G., Reznick, R., MacRae, H., Murnaghan, J., Hutchison, C., et al. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg, 84*(2), 273-278.

Matthews, A. H., Abdelrahman, T., Powell, A. G., & Lewis, W. G. (2016). Surgical Education's 100 Most Cited Articles: A Bibliometric Analysis. *J Surg Educ, 73*(5), 919-929. https://doi.org/10.1016/j.jsurg.2016.05.011

Meyers, W. C. B., Gene D; et al. (1991). A prospective analysis of 1518 laparoscopic cholecystectomies. The Southern Surgeons Club. *N Engl J Med, 324*(16), 1073-1078. https://doi.org/10.1056/NEJM199104183241601

Mitsouras, D., Liacouras, P., Imanzadeh, A., Giannopoulos, A. A., Cai, T., Kumamaru, K. K., et al. (2015). Medical 3D Printing for the Radiologist. *Radiographics, 35*(7), 1965-1988. https://doi.org/10.1148/rg.2015140320

Munz, Y., Kumar, B. D., Moorthy, K., Bann, S., & Darzi, A. (2004). Laparoscopic virtual reality and box trainers: is one superior to the other? *Surg Endosc, 18*(3), 485-494. https://doi.org/10.1007/s00464-003-9043-7

Nagendran, M., Gurusamy, K. S., Aggarwal, R., Loizidou, M., & Davidson, B. R. (2013). Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev*(8), CD006575. https://doi.org/10.1002/14651858.CD006575.pub3

Nezhat, F. (2003). Triumphs and controversies in laparoscopy: the past, the present, and the future. *JSLS, 7*(1), 1-5.

Nicholas, R., Humm, G., MacLeod, K., Bathla, S., Horgan, A., Nally, D. M., et al. (2019). Simulation In Surgical Training: Prospective Cohort Study Of Access, Attitudes And Experiences Of Surgical Trainees In The Uk And Ireland. *Int J Surg*. https://doi.org/10.1016/j.ijsu.2019.04.004

Niitsu, H., Hirabayashi, N., Yoshimitsu, M., Mimura, T., Taomoto, J., Sugiyama, Y., et al. (2013). Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today, 43*(3), 271-275. https://doi.org/10.1007/s00595-012-0313-7

Okrainec, A., Soper, N. J., Swanstrom, L. L., & Fried, G. M. (2011). Trends and results of the first 5 years of Fundamentals of Laparoscopic Surgery (FLS) certification testing. *Surgical endoscopy, 25*(4), 1192-1198. https://doi.org/10.1007/s00464-010-1343-0

Oropesa, I., Escamirosa, F. P., Sanchez-Margallo, J. A., Enciso, S., Rodriguez-Vila, B., Martinez, A. M., et al. (2018). Interpretation of motion analysis of laparoscopic instruments based on principal component analysis in box trainer settings. *Surg Endosc*. https://doi.org/10.1007/s00464-018-6022-6

Oropesa, I., Sanchez-Gonzalez, P., Lamata, P., Chmarra, M. K., Pagador, J. B., Sanchez-Margallo, J. A., et al. (2011). Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. *The Journal of surgical research, 171*(1), e81-95. https://doi.org/10.1016/j.jss.2011.06.034

Overtoom, E. M., Horeman, T., Jansen, F. W., Dankelman, J., & Schreuder, H. W. R. (2018). Haptic Feedback, Force Feedback, and Force-Sensing in Simulation Training for Laparoscopy: A Systematic Overview. *J Surg Educ*. https://doi.org/10.1016/j.jsurg.2018.06.008

Panait, L., Akkary, E., Bell, R. L., Roberts, K. E., Dudrick, S. J., & Duffy, A. J. (2009). The role of haptic feedback in laparoscopic simulation training. *J Surg Res, 156*(2), 312-316. https://doi.org/10.1016/j.jss.2009.04.018

Perissat, J. (1999). Laparoscopic surgery: A pioneer's point of view. *World J Surg, 23*(8), 863-868. https://doi.org/10.1007/s002689900590

Peters, J. H., Fried, G. M., Swanstrom, L. L., Soper, N. J., Sillin, L. F., Schirmer, B., et al. (2004). Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery, 135*(1), 21-27. https://doi.org/10.1016/S0039-6060(03)00156-9

Picod, G., Jambon, A. C., Vinatier, D., & Dubois, P. (2005). What can the operator actually feel when performing a laparoscopy? *Surg Endosc, 19*(1), 95-100. https://doi.org/10.1007/s00464-003-9330-3

Polavarapu, H. V., Kulaylat, A. N., Sun, S., & Hamed, O. H. (2013). 100 years of surgical education: the past, present, and future. *Bull Am Coll Surg, 98*(7), 22-27.

Prentice, R. (2012). *Bodies in formation: An ethnography of anatomy and surgery education*. Durham, USA: Duke University Press.

RACS. (2019). *Competencies - ROYAL AUSTRALASIAN COLLEGE OF SURGEONS*. Retrieved 29.05.2019, from https://www.surgeons.org/becoming-a-surgeon/surgical-education-training/competencies/

Rashid, P. (2017). Surgical education and adult learning: Integrating theory into practice. *F1000Res, 6*, 143. https://doi.org/10.12688/f1000research.10870.1

RCS. (2019). *Competencies - The Royal College of Surgeons of England*. Retrieved 29.05.2018, from https://www.rcseng.ac.uk/careers-in-surgery/trainees/foundation-and-core-trainees/competencies/

Rengier, F., Mehndiratta, A., von Tengg-Kobligk, H., Zechmann, C. M., Unterhinninghofen, R., Kauczor, H. U., et al. (2010). 3D printing based on imaging data: review of medical applications. *Int J Comput Assist Radiol Surg, 5*(4), 335-341. https://doi.org/10.1007/s11548-010-0476-x

Richards, C., Rosen, J., Hannaford, B., Pellegrini, C., & Sinanan, M. (2000). Skills evaluation in minimally invasive surgery using force/torque signatures. *Surg Endosc, 14*(9), 791-798. https://doi.org/10.1007/s004640000230

Richling, B. (2006). History of Endovascular Surgery: Personal Accounts of the Evolution. *Neurosurgery 59*(suppl 5), s3-30-s33-38.

Robinson, T. N., & Stiegmann, G. V. (2004). Minimally invasive surgery. *Endoscopy, 36*(1), 48-51. https://doi.org/10.1055/s-2004-814113

Rodrigues, S. P., Horeman, T., Dankelman, J., van den Dobbelsteen, J. J., & Jansen, F. W. (2012). Suturing intraabdominal organs: when do we cause tissue damage? *Surg Endosc, 26*(4), 1005-1009. https://doi.org/10.1007/s00464-011-1986-5

Rodrigues, S. P., Horeman, T., Dankelman, J., van den Dobbelsteen, J. J., & Jansen, F. W. (2015). Tying different knots: what forces do we use? *Surgical Endoscopy, 29*(7), 1982-1989. https://doi.org/10.1007/s00464-014-3898-7

Royal, K. D. (2017). Four tenets of modern validity theory for medical education assessment and evaluation. *Adv Med Educ Pract, 8*, 567-570. https://doi.org/10.2147/AMEP.S139492

Ruthenbeck, G., & Reynolds, K. (2013). Virtual reality surgical simulator software development tools. *Journal of Simulation, 7*(2), 101-108. https://doi.org/https://doi.org/10.1057/jos.2012.22

Sadideen, H., Hamaoui, K., Saadeddin, M., & Kneebone, R. (2012a). Simulators and the simulation environment: getting the balance right in simulation-based surgical education. *Int J Surg, 10*(9), 458-462. https://doi.org/10.1016/j.ijsu.2012.08.010

Sadideen, H., & Kneebone, R. (2012b). Practical skills teaching in contemporary surgical education: how can educational theory be applied to promote effective learning? *Am J Surg, 204*(3), 396-401. https://doi.org/10.1016/j.amjsurg.2011.12.020

Salkini, M. W., Doarn, C. R., Kiehl, N., Broderick, T. J., Donovan, J. F., & Gaitonde, K. (2010). The role of haptic feedback in laparoscopic training using the LapMentor II. *J Endourol, 24*(1), 99-102. https://doi.org/10.1089/end.2009.0307

Satava, R. M. (1993). Virtual reality surgical simulator. The first steps. *Surg Endosc, 7*(3), 203-205.

Satava, R. M. (2001). Surgical education and surgical simulation. *World J Surg, 25*(11), 1484-1489.

Seymour, N. E. (2008). VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World J Surg, 32*(2), 182-188. https://doi.org/10.1007/s00268-007-9307-9

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., et al. (2002). Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg, 236*(4), 458-463; discussion 463-454. https://doi.org/10.1097/01.SLA.0000028969.51489.B4

Shanafelt, T. D., Balch, C. M., Bechamps, G., Russell, T., Dyrbye, L., Satele, D., et al. (2010). Burnout and medical errors among American surgeons. *Ann Surg, 251*(6), 995-1000. https://doi.org/10.1097/SLA.0b013e3181bfdab3

Smith, R. (1998). All changed, changed utterly. British medicine will be transformed by the Bristol case. *BMJ, 316*(7149), 1917-1918. https://doi.org/10.1136/bmj.316.7149.1917

Stefanidis, D., Arora, S., Parrack, D. M., Hamad, G. G., Capella, J., Grantcharov, T., et al. (2012). Research priorities in surgical simulation for the 21st century. *American journal of surgery, 203*(1), 49-53. https://doi.org/10.1016/j.amjsurg.2011.05.008

Stefanidis, D., Korndorffer Jr., J. R., & Sweet, R. (Eds.). (2019). *Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties*. Springer Nature Switzerland AG 2019: Springer, Cham.

Strom, P., Hedman, L., Sarna, L., Kjellin, A., Wredmark, T., & Fellander-Tsai, L. (2006). Early exposure to haptic feedback enhances performance in surgical simulator training: a prospective randomized crossover study in surgical residents. *Surg Endosc, 20*(9), 1383-1388. https://doi.org/10.1007/s00464-005-0545-3

Subramonian, K., DeSylva, S., Bishai, P., Thompson, P., & Muir, G. (2004). Acquiring surgical skills: a comparative study of open versus laparoscopic surgery. *Eur Urol, 45*(3), 346-351; author reply 351. https://doi.org/10.1016/j.eururo.2003.09.021

Sweet, R. M., Hananel, D., & Lawrenz, F. (2010). A unified approach to validation, reliability, and education study design for surgical technical skills training. *Arch Surg, 145*(2), 197-201. https://doi.org/10.1001/archsurg.2009.266

Szasz, P., Louridas, M., Harris, K. A., Aggarwal, R., & Grantcharov, T. P. (2015). Assessing Technical Competence in Surgical Trainees: A Systematic Review. *Ann Surg, 261*(6), 1046-1055. https://doi.org/10.1097/SLA.0000000000000866

Takayasu, K., Yoshida, K., Kinoshita, H., Yoshimoto, S., Oshiro, O., & Matsuda, T. (2018). Analysis of the tractive force pattern on a knot by force measurement during laparoscopic knot tying. *The American Journal of Surgery, 216*(2), 314-318. https://doi.org/10.1016/j.amjsurg.2017.07.009

Tavakol, M., Mohagheghi, M. A., & Dennick, R. (2008). Assessing the skills of surgical residents using simulation. *J Surg Educ, 65*(2), 77-83. https://doi.org/10.1016/j.jsurg.2007.11.003

Thinggaard, E., Bjerrum, F., Strandbygaard, J., Gogenur, I., & Konge, L. (2016). Ensuring Competency of Novice Laparoscopic Surgeons-Exploring Standard Setting Methods and their Consequences. *J Surg Educ, 73*(6), 986-991. https://doi.org/10.1016/j.jsurg.2016.05.008

Tholey, G., Desai, J. P., & Castellanos, A. E. (2005). Force feedback plays a significant role in minimally invasive surgery: results and analysis. *Ann Surg, 241*(1), 102-109. https://doi.org/10.1097/01.sla.0000149301.60553.1e

Thompson, J. R., Leonard, A. C., Doarn, C. R., Roesch, M. J., & Broderick, T. J. (2011). Limited value of haptics in virtual reality laparoscopic cholecystectomy training. *Surg Endosc, 25*(4), 1107-1114. https://doi.org/10.1007/s00464-010-1325-2

Treasure, T. (1998). Lessons from the Bristol case. More openness--on risks and on individual surgeons' performance. *BMJ, 316*(7146), 1685-1686.

Trejos, A. L., Patel, R. V., Malthaner, R. A., & Schlachta, C. M. (2014). Development of force-based metrics for skills assessment in minimally invasive surgery. *Surg Endosc, 28*(7), 2106-2119. https://doi.org/10.1007/s00464-014-3442-9

Underwood, W., 3rd, Fletcher, K. E., & Chen, S. L. (2003). How will limiting resident work hours impact surgical training? *Curr Surg, 60*(6), 578-584. https://doi.org/10.1016/S0149-7944(03)00156-9

Van Bruwaene, S., Schijven, M. P., Napolitano, D., De Win, G., & Miserez, M. (2015). Porcine cadaver organ or virtual-reality simulation training for laparoscopic cholecystectomy: a randomized, controlled trial. *J Surg Educ, 72*(3), 483-490. https://doi.org/10.1016/j.jsurg.2014.11.015

van den Dobbelsteen, J. J., Schooleman, A., & Dankelman, J. (2007). Friction dynamics of trocars. *Surg Endosc, 21*(8), 1338-1343. https://doi.org/10.1007/s00464-006-9105-8

van der Meijden, O. A., & Schijven, M. P. (2009). The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review. *Surg Endosc, 23*(6), 1180-1190. https://doi.org/10.1007/s00464-008-0298-x

van Dongen, K. W., Ahlberg, G., Bonavina, L., Carter, F. J., Grantcharov, T. P., Hyltander, A., et al. (2011). European consensus on a competency-based virtual reality training program for basic endoscopic surgical psychomotor skills. *Surg Endosc, 25*(1), 166-171. https://doi.org/10.1007/s00464-010-1151-6

van Hove, P. D., Tuijthof, G. J., Verdaasdonk, E. G., Stassen, L. P., & Dankelman, J. (2010). Objective assessment of technical surgical skills. *Br J Surg, 97*(7), 972-987. https://doi.org/10.1002/bjs.7115

Våpenstad, C., & Buzink, S. N. (2013a). Procedural virtual reality simulation in minimally invasive surgery. *Surg Endosc, 27*(2), 364-377. https://doi.org/10.1007/s00464-012-2503-1

Våpenstad, C., Hofstad, E. F., Bo, L. E., Chmarra, M. K., Kuhry, E., Johnsen, G., et al. (2013b). Limitations of haptic feedback devices on construct validity of the LapSim(R) virtual reality simulator. *Surg Endosc, 27*(4), 1386-1396. https://doi.org/10.1007/s00464-012-2621-9

Våpenstad, C., Hofstad, E. F., Bø, L. E., Kuhry, E., Johnsen, G., Mårvik, R., et al. (2017). Lack of transfer of skills after virtual reality simulator training with haptic feedback. *Minim Invasive Ther Allied Technol, 26*(6), 346-354. https://doi.org/10.1080/13645706.2017.1319866

Våpenstad, C., Hofstad, E. F., Eivind Bernstein, T., Aadahl, P., Johnsen, G., & Mårvik, R. (2019). Optimal timing of assessment tasks depending on experience level of surgical trainees. *Minim Invasive Ther Allied Technol*, 1-9. https://doi.org/10.1080/13645706.2019.1612441

Våpenstad, C., Hofstad, E. F., Lango, T., Marvik, R., & Chmarra, M. K. (2013c). Perceiving haptic feedback in virtual reality simulators. *Surgical Endoscopy, 27*(7), 2391-2397. https://doi.org/10.1007/s00464-012-2745-y

Vassiliou, M. C., & Feldman, L. S. (2011). Objective assessment, selection, and certification in surgery. *Surg Oncol, 20*(3), 140-145. https://doi.org/10.1016/j.suronc.2010.09.004

Vassiliou, M. C., Feldman, L. S., Andrew, C. G., Bergman, S., Leffondre, K., Stanbridge, D., et al. (2005). A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg, 190*(1), 107-113. https://doi.org/10.1016/j.amjsurg.2005.04.004

Volodos, N. L. (2015). The 30th Anniversary of the First Clinical Application of Endovascular Stent-grafting. *Eur J Vasc Endovasc Surg, 49*(5), 495-497. https://doi.org/10.1016/j.ejvs.2015.02.012

Wanzel, K. R., Ward, M., & Reznick, R. K. (2002). Teaching the surgical craft: From selection to certification. *Curr Probl Surg, 39*(6), 573-659. https://doi.org/10.1067/mog.2002.123481

Waran, V., Narayanan, V., Karuppiah, R., Thambynayagam, H. C., Muthusamy, K. A., Rahman, Z. A., et al. (2015). Neurosurgical endoscopic training via a realistic 3-dimensional model with pathology. *Simul Healthc, 10*(1), 43-48. https://doi.org/10.1097/SIH.0000000000000060

Watson, R. A. (2014). A low-cost surgical application of additive fabrication. *J Surg Educ, 71*(1), 14-17. https://doi.org/10.1016/j.jsurg.2013.10.012

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*: Cambridge university press.

Westebring-van der Putten, E. P., Goossens, R. H., Jakimowicz, J. J., & Dankelman, J. (2008). Haptics in minimally invasive surgery--a review. *Minim Invasive Ther Allied Technol, 17*(1), 3-16. https://doi.org/10.1080/13645700701820242

Westebring-van der Putten, E. P., van den Dobbelsteen, J. J., Goossens, R. H., Jakimowicz, J. J., & Dankelman, J. (2009). Effect of laparoscopic grasper force transmission ratio on grasp control. *Surgical endoscopy, 23*(4), 818-824. https://doi.org/10.1007/s00464-008-0107-6

Willaert, W. I., Aggarwal, R., Van Herzeele, I., Cheshire, N. J., & Vermassen, F. E. (2012). Recent advancements in medical simulation: patient-specific virtual reality simulation. *World J Surg, 36*(7), 1703-1712. https://doi.org/10.1007/s00268-012-1489-0

Willaert, W. I., Aggarwal, R., Van Herzeele, I., O'Donoghue, K., Gaines, P. A., Darzi, A. W., et al. (2011). Patient-specific endovascular simulation influences interventionalists performing carotid artery stenting procedures. *Eur J Vasc Endovasc Surg, 41*(4), 492-500. https://doi.org/10.1016/j.ejvs.2010.12.013

Windsor, J. A. (2009). Role of simulation in surgical education and training. *ANZ journal of surgery, 79*(3), 127-132. https://doi.org/10.1111/j.1445-2197.2008.04829.x

Woodrum, D. T., Andreatta, P. B., Yellamanchilli, R. K., Feryus, L., Gauger, P. G., & Minter, R. M. (2006). Construct validity of the LapSim laparoscopic surgical simulator. *Am J Surg, 191*(1), 28-32. https://doi.org/10.1016/j.amjsurg.2005.10.018

Yiannakopoulou, E., Nikiteas, N., Perrea, D., & Tsigris, C. (2015). Virtual reality simulators and training in laparoscopic surgery. *Int J Surg, 13*, 60-64. https://doi.org/10.1016/j.ijsu.2014.11.014

Zhou, M., Tse, S., Derevianko, A., Jones, D. B., Schwaitzberg, S. D., & Cao, C. G. (2012). Effect of haptic feedback in laparoscopic surgery skill acquisition. *Surgical endoscopy, 26*(4), 1128-1134. https://doi.org/10.1007/s00464-011-2011-8

Paper 1

Perceiving Haptic Feedback in Virtual Reality Simulators

Våpenstad C, Hofstad EF, Langø T, Mårvik R, Chmarra MK

Surgical Endoscopy 27(7), 2391-2397, 2013

Paper 2

Procedural Virtual Reality Simulation in Minimally Invasive Surgery

Våpenstad C, Buzink SN

Surgical Endoscopy 27(2), 364-377, 2013

Paper 2

Online Supplementary Material

Is not included due to copyright available at https://doi.org/10.1007/s00464-012-2503-1

Paper 3

Limitations of Haptic Feedback Devices on Construct Validity of the LapSim® Virtual Reality Simulator

Våpenstad C, Hofstad EF, Bø LE, Chmarra MK, Kuhry E, Johnsen G, Mårvik R, Langø T Surgical Endoscopy 27(4), 1386-1396, 2013

Paper 4

Lack of Transfer of Skills after Virtual Reality Simulator Training with Haptic Feedback

Våpenstad C, Hofstad EF, Bø LE, Kuhry E, Johnsen G, Mårvik R, Langø T, Hernes TN Minimally Invasive Therapy & Allied Technologies, 26(6), 346-354, 2017

Paper 5

Optimal Timing of Assessment Tasks Depending on Experience Level of Surgical Trainees

Våpenstad C, Hofstad EF, Bernstein TE, Aadahl P, Johnsen G, Mårvik R

Minimally Invasive Therapy & Allied Technologies, 1-9, 2019

Doctoral theses at NTNU, 2019:323

Cecilie Våpenstad

Simulation-based Training and Assessment in Minimally Invasive Surgery

Exploration and Validation

Doctoral Thesis

Cecilie Våpenstad

Doctoral theses at NTNU, 2019:323

NTNU
Norwegian University of
Science and Technology
Faculty of Medicine and Health Sciences
Department of Circulation
and Medical Imaging

ISBN 978-82-326-4242-7 (printed version)
ISBN 978-82-326-4243-4 (electronic version)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU