“Sign them in” -
Proficiency - based surgical education:
An advantage for patients, doctors and employers

Thesis by
Jeanne Mette Goderstad

SØRLANDET HOSPITAL

Oslo
University Hospital

UiO : Faculty of Medicine
University of Oslo

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-348-0151-8

Print production: Graphics Center, University of Oslo.

We can be serious and still have fun.

0. Acknowledgements

Fulfilling a PhD is achieved with help and support. This PhD had not been possible without my main supervisor, Professor Marit Lieng. I am very grateful to her. Thank you for your patience, and for not giving up on me, as the years have passed.

I am also very grateful to my other supervisors, Professor Erik Fosse and Professor Leiv Sandvik. You are both inspiring. I have learned so much through our discussions. You made a difference.

I thank Professor Trond Buanes. He introduced me to Professor Neal Seymour at Ullevål University Hospital in 2004. This meeting gave inspiration and started my interest for training in Minimally Invasive Surgery. Further I want to thank Professor Buanes for recommending me for this PhD programme.

I thank Bjørn Busund, my boss at Ullevål University Hospital for giving me the possibility to start the ph.d programme by providing time and financial support.

Also thanks to Torunn Fetveit, one of my mentors for telling me about the first World Congress of Surgical Training. At this congress I met the pioneers within surgical training.

I am grateful to Jelena Kisic and Anton Langebrekke for inspiration, expertise and valuable contribution by assessing the video recorded laparoscopic supracervical hysterectomy procedures.

Thanks to my colleagues and participants who have contributed with their time, effort and enthusiasm.

I thank my friends for togetherness and conversations about surgical training, work, family, social and political issues.

To my dear mother, father and brother, for a good upbringing, always being there for me with help, support and constructive feedback.

My deepest gratitude to my husband Olav Anders, for support and patience.

Our daughters Olea and Tekla, I thank you for reminding me about the most important thing in life.

1. Table of contents

2. Abbreviations

AAGL: American Association of Gynecological Laparoscopists

CAT-LSH: Competence Assessment Tool- Laparoscopic supracervical hysterectomy

CME: Continuous Medical Education

CT: Computerized tomography

CUSUM: Cumulative Summation

ECRES: Certification for Reproductive Endoscopic Surgery

ESGE: European Society for Gynaecological Endoscopy

FLS: Fundamentals of Laparoscopic Surgery

GESEA: The Gynaecological Endoscopic Surgical Education

GOALS: Global Operative Assessment in Laparoscopic Surgery

GOALS-GH: Global Operative Assessment in Laparoscopic Surgery-

MRI: Magnetic resonance imaging

NFOG: Nordic Federation of Societies of Obstetrics and Gynecology.

NSGE: Nordic Society of Gynaecological Endoscopy

OSATS: The Objective Structured Assessment of Technical Skills

O-SCORE: Ottawa Surgical Competency Operating Room Evaluation

OSCEs: Objective Structured Clinical Examinations

VR: Virtual reality

3. List of papers

1. Goderstad JM, Sandvik L, Fosse E, Lieng M. Assessment of Surgical Competence: Development and Validation of Rating Scales Used for Laparoscopic Supracervical Hysterectomy. J. Surg Ed 73:600-608. 2016.

2. Goderstad JM, Sandvik L, Fosse E, Lieng M. Development and validation of a general and easy assessable scoring system for laparoscopic skills using a virtual reality simulator. Eur J Obstet Gynecol Reprod Biol X. 2019 aug 13;4:100092.doi:10.1016/j.eurox.2019.100092.eCollection 2019 Oct.

3. Goderstad JM, Sandvik L, Fosse E, Lieng M. Development and validation of a curriculum for laparoscopic supracervical hysterectomy. Facts Viewes and Vis OBgyn.2020 Aug 5;12(2):83-90

The contribution of Jeanne Mette Goderstad was:

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

Paper 2: Major role in design of the study, data collection, analysis and writing the paper

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

4. Summaries

4.1 Summary in Norwegian.

Vi forventer å bli behandlet av kompetente leger når vi trenger medisinsk og kirurgisk behandling. Denne kompetansen er en kombinasjon av faktorer, som blant annet kunnskap, tekniske ferdigheter, besluttsomhet, kommunikasjon og lederegenskaper. Ferdighetene kan læres og tilbakemelding er en avgjørende faktor for å sikre at ferdighetene oppnås.
Kirurgi krever åpninger i kroppen som muliggjør undersøkelse, korrigering eller fjerning av organer, strukturer og vev. Ved minimal invasiv operasjonsteknikk gjennomføres den kirurgiske prosedyren gjennom små insisjoner. Laparoskopi er en minimal invasiv operasjonsteknikk som gir tilgang til bukhulen og bekkenet. I 1995 utviklet gynekologen Kurt Semm en læringsmodell for trening i laparoskopisk kirurgi. Det var en boks med hull for innføring av instrumenter og et kamera. På denne måten kunne man manipulere ulike objekter inne i boksen, på samme måte som man gjør det i buken til en pasient under en operasjon. Med dette konseptet kunne leger skaffe seg kirurgiske ferdigheter før de opererte pasienter. I løpet av 1990 tallet utviklet teknologien seg, og det var mulig å ta i bruk simulatorer og virtuell virkelighet for trening. Vi kan velge kirurgiske prosedyrer eller øvelser for basale ferdigheter fra en meny på simulatoren. Når øvelsene er gjennomført, får vi umiddelbart en rapport med objektive og standardiserte data på utførelsen. Flere forskningsgrupper har gjennomført randomiserte studier der målet var å evaluere om ferdighetstrening på simulator gjør deg til en bedre operatør på operasjonsstuen. Ikke overraskende viste det seg at deltagere som gjennomførte strukturert trening før de opererte pasienter, gjennomførte den kirurgiske behandlingen raskere og med større presisjon en deltagere som ikke hadde gjennomført treningen på forhånd.

Hysterektomi (fjernelse av livmoren) er en vanlig gynekologisk prosedyre. Det gjennomføres om lag 4500-5000 hysterektomier i Norge per år. Internasjonale retningslinjer anbefaler minimal invasiv operasjonsteknikk når forholdene ligger til rette for dette.
Målet for studiene som inngår i avhandlingen var å utvikle og evaluere et treningsprogram for laparoskopisk supracervical hysterektomi. Dette treningsprogrammet skulle være praktisk og gjennomførbart for utdanningen av gynekologer i Norge.

I studie1 samarbeidet vi med en ekspertgruppe om utformingen av et prosedyrespesifikt evalueringsskjema for laparoskopisk supracervical hysterectomi, CAT- LSH. Vi sammenlignet dette med et generelt evalueringsskjema for laparoskopisk kirurgi, Global Operative Assessment of Laparoscopic Skills, GOALS. Vi inkluderte 37 laparoskopiske supracervicale hysterectomier utført av leger i spesialisering og gynekologer med ulik kirurgisk kompetanse. Alle prosedyrer ble tatt opp på video og evaluert med de to evalueringsskjemanene av to blindede observatører. Assistentene under inngrepene evaluerte i tillegg utførelsen umiddelbart etter operasjonens slutt. Vi fant en statistisk signifikant forskjell på gjennomsnittsskåren for de uerfarne, de med noe erfaring og de erfarne deltagerne ved bruk av begge evalueringsskjema. De erfarne fikk høyest skår, de med noe erfaring lavere skår og de uerfarne lavest skår. Samsvar mellom observatørene på skåringen med evaluerinsskjemaene var meget god. Vi konkluderte med at evaluering av kirurgisk kompetanse under en laparoskopisk supracervikal hysterektomi kan gjøres med Competence Assessment Tool for Laparoscopic Supracervikal Hysterectomy og Global Operative Assessment of Laparoscopic Skills.

Kompetansebasert utdanning tar utgangspunkt i oppnåelse av ferdigheter. Kravene er definert utfra ferdighetene til erfarne kolleger. I studie 2 utarbeidet og evaluerte vi et skåringssystem på fem øvelser på en laparoskopisimulator. Vi inkluderte 30 gynekologer og leger i spesialisering. Deltagerne ble kategorisert inn i tre grupper utfra kirurgisk erfaring; uerfarne, noe erfarne og erfarne. Alle studiedeltakerne utførte 10 repetisjoner av hver øvelse. Øvelse 1,2 og 3 var basale ferdighetsøvelser, øvelse 4 en salpingektomi (fjernelse av eggleder) og øvelse 5 en laparoskopisk supracervikal hysterektomi. Vi registrerte tid, feil og instrumentbevegelser i gjennomføringen av alle øvelsene. For å definere krav til ferdighet på hver øvelse, brukte vi gjennomsnittet av de erfarne deltagernes resultater på de fire siste repetisjonene. Lik eller bedre utførelse av dette gjennomsnittet gav en poengsum på 2. En lavere skår tilsvarende et standardavvik gav 1 poeng og alle resultater lavere gav 0 poeng. Det var signifikante forskjeller i gjennomsnittlig totalskår mellom de uerfarne og de erfarne for øvelse 1, 2, 4 og 5. Det var statistisk signifikant forskjell i totalskår mellom de noe erfarne og de erfarne deltagerne for øvelse 1 og 3. Det var ingen statistisk signifikant forskjell mellom de uerfarne og de noe erfarne på noen av øvelsene. Vi konkluderte med at skåringssystemet kan brukes i kompetansebasert utdanning.

Innen opplæring i laparoskopi er evaluering og sertifisering ansett som en gullstandard. På tross av dette er det mange sykehus som ikke benytter tilgjengelige opplæringsprogrammer for laparoskopisk kirurgi. I studie 3 designet og evaluerte vi et opplæringsprogram for laparoskopisk supracervikal hysterektomi. Første del var teoretisk og besto av en validert multipel-choice test. Del to var praktisk og besto av fem øvelser på en laparoskopisimulator. Når deltagerne oppnådde ferdighetskravet vi hadde definert på forhånd, var de klare for del 3 som var en laparoskopisk supracervical hysterectomi. Alle prosedyrer ble tatt opp på video og evaluert av to blindede observatører. Vi inkluderte 12 uerfarne leger i spesialisering som deltagere. Alle gjennomførte opplæringsprogrammet. Gjennomsnittstiden på treningsperioden (del 1 og 2) var 57.0 dager (SD 26.0). Deltagerne brukte i gjennomsnitt 173.0 minutter (SD 49.0) på simulatoren for å nå ferdighetstravet for alle øvelsene. De uerfarne deltagerne som gjennomførte opplæringsprogrammet, hadde en statistisk signifikant bedre skår på utførelsen av den laparoskopiske supracervikale hysterectomien enn den uerfarne gruppen leger i spesialisering som vi presenterte i artikkel 1.

Med ferdighetsbasert opplæring får pasientene leger med den tilstrekkelige kirurgiske kompetanse, leger i spesialisering får et forutsigbart opplæringsprogram som gjør dem forberedt for den kirurgiske behandlingen de skal utføre, og arbeidsgivere får dokumentasjon på at ansatte har den ferdigheten som er ansett nødvendig for den jobben de er satt til å gjøre

4.2. Summary in English

We expect to be treated by competent doctors when we need medical help and surgery. This competence is a combination of knowledge, technical skills, decision making, communication, and leadership. Competencies can be learned and feedback is decisive to achieve good learning.

Education for doctors is lifelong. It is however necessary to ensure that certain standards are reached at certain points in a medical doctor's career.

Surgery requires incisions into the body. Minimally invasive surgery, is a technique that requires small incisions where we can introduce instruments and examine, correct or remove organs, structures and tissue. Laparoscopy is a minimally invasive surgical technique that gives the surgeon access to the abdomen. In 1985, the gynaecologist, Kurt Semm designed a didactic tool for training in laparoscopic surgery. It was a box with holes for instruments, a camera and different objects that could be manipulated by laparoscopic instruments inside the box in the same way as in abdomen in humans. With this concept, doctors got the possibility to practise, and acquire surgical skills outside the operation theatre. During the 1990`s, technology advanced and it was possible to use virtual reality simulators. According to the Healthcare Simulation Dictionary, virtual simulation is "the recreation of reality depicted on a computer screen". The trainee can select different tasks, like surgical procedure or tasks for basic skills. When the task is completed, an automated report provides feedback. The feedback is objective and standardized.

Several research groups have evaluated skills transfer from virtual reality simulators to operations on humans by randomised controlled trials. The intervention groups, the participants with structured training before entering the operation theatre, perform surgery faster and more safely than participants without training.

Hysterectomy is a common gynaecological procedure. In Norway we perform around 4500-5000 hysterectomies per year. Guidelines recommend minimally invasive techniques when possible. The aim of the studies included in the Thesis is to evaluate the feasibility and effect of training on a virtual simulator before performing laparoscopic hysterectomy.

Assessment is essential for constructive feedback. In study1, we collaborated with an expert group and developed a procedure-specific rating scale for laparoscopic supracervical hysterectomy (CAT- LSH) and compared it with a general rating scale in laparoscopic

surgery, Global Operative Assessment of Laparoscopic Skills, GOALS. We included 37 laparoscopic supracervical procedures performed by gynaecologists with different levels of surgical competence. All procedures were video-recorded and assessed by two blinded observers and the assistant surgeons. There were statistically significant difference between mean score for the inexperienced, intermediate experienced and experienced participants for both rating scales and the interrater reliability was high. We concluded that assessment of surgical competence during LSH is feasible with the Competence Assessment Tool for Laparoscopic Supracervical Hysterectomy and the rating scale Global Operative Assessment of Laparoscopic Skills.

Proficiency-based education is based on experts performing a task or procedure to define the training goals. In study 2, we developed and validated a scoring system for laparoscopic skills for five tasks on a virtual reality simulator. We included 30 trainees and gynaecologists. The participants were categorized as inexperienced, intermediate experienced and experienced. They performed ten repetitions of the different tasks. Task 1, 2 and 3 were tasks in basic skills, task 4 a salpingectomy, and task 5 a laparoscopic supracervical hysterectomy. We registered time, error parameters, and economy of movements measured by the simulator. We used the results of the 4 last repetitions performed by the experienced as base for a scoring system. Performance equal to, and higher than, the mean score gave 2 points. A decrease of 1 SD from the mean gave 1 point, and every score below gave 0 points.
There was a statistical significant difference in total score between the inexperienced and the experienced group in tasks 1, 3, 4 and 5. There was a statistical significant difference in total score between the intermediate experienced and the experienced group in tasks 1 and 3. There was no statistical significant difference between the inexperienced and the intermediate experienced group for any task. We concluded that the scoring system can be used in proficiency-based assessment.

Assessment and certification of skills within minimal invasive surgery are considered a gold standard for assuring that a surgeon has acquired and retained a certain level of knowledge and skills. Hospitals training gynaecologists rarely use available educational programmes for endoscopic surgery. In study 3, we designed and evaluated a three-step curriculum for laparoscopic supracervical hysterectomy. Step 1, the theoretical part, was a validated multiple-choice test. Step 2 was a practical part consisting of five tasks on a virtual simulator. When the participants reached the pre-set proficiency levels on the simulator they moved on to step 3, the laparoscopic supracervical hysterectomy. The procedures were video recorded, and assessed with the procedure specific competence assessment tool for laparoscopic supracervical hysterectomy and the general rating scale in laparoscopic surgery, Global Operative Assessment of Laparoscopic Skills by two blinded observers and the assistant surgeons. We included 12 participants and all of them completed the curriculum. The mean duration of the training period (step 1 and 2) was 57.0 days (SD 26.0). The participants spent a mean of 173.0 min (SD 49.0) on the simulator to reach the pre-set proficiency level for all tasks. The inexperienced participants who completed the curriculum had a statistically significantly better score on the video evaluation of the performance of the laparoscopic supracervical hysterectomy, than the inexperienced group without the structured training that we presented in paper 1.

Safer surgeons faster, summarizes the benefits with virtual reality simulator training. Sign them in, summarizes the way to achieve it.
With proficiency-based education, the patients get doctors with surgical skills, trainees get a predictable training program and are prepared for surgery, and employers know that employees have the competence necessary for the job they are supposed to do.

4.3. Thesis at a glance

	Aims	**Study design**	**Results**	**Conclusion**
1	Develop a procedure-specific rating scale for laparoscopic supracervical hysterectomy (CAT-LSH), and compare the construct validity and interrater reliability with a general rating scale in laparoscopic surgery (GOALS)	Prospective, inter-observer study. Comparing a general and a procedure specific rating scale for LSH. Gynaecologists with different surgical experience performed a LSH that was video recorded and evaluated by the surgical assistant and two blinded observers. The study was performed in Jan 2013-Sept 2013	37 procedures were included. Inexperienced/intermediate experienced surgeons**:** Evaluation by the assistant surgeon and the blinded observers showed statistically significant differences between the two groups: GOALS, p<0.001 and p=0.001, CAT-LSH p< 0.002 and p=0.006. Intermediate experienced/ experienced surgeons: Difference between the two groups of surgeons using GOALS was statistically significant, p=0.002. When assessed by the assistant surgeon CAT-LSH showed statistically significant differences, both when assessed by the assistant surgeon, p<0.001, and the blinded observers, p=0.001. Interclass correlation coefficient Blinded observers GOALS 0.74, CAT-LSH 0.85. Interclass correlation coefficient Blinded observers / surgical assistants GOALS: 0.71, CAT-LSH: 0.75.	There were significant differences between the proficiency groups. The CAT-LSH and GOALS appeared to have construct validity and high interrater reliability**.**
2	Develop and validate a scoring system for laparoscopic skills.	A prospective, longitudinal cohort study including 30 gynaecologists and trainees at three hospitals. The participants performed ten repetitions of five tasks on a VR simulator. The study was performed in Sept 2013-May 2014.	There was a statistically significant difference of in total score when comparing the inexperienced and the experienced group in task 1: 3.4 (SD 0.6) vs 5.1 (SD 1.1), p=0.01, task 3: 1.7 (SD 0.7) vs 2.8 (SD 0.5), p<0.01, task 4, 3.6 min (SD 1.4 min) vs 2.3 min (SD 1.0 min), p=0.03, and task 5, 3.2 (SD 1.5) vs 5.3 (SD 1.8), p=0.01. There was no statistically significant difference in total score between the inexperienced and the intermediate group in any of the tasks. When comparing the intermediate and the experienced group, there was a statistically significant difference in task 1: 3.4 (SD 0.6) vs 5.1 (SD 1.1), p=0.01 and task 3, 1.7 (SD 0.7) vs 1.9 (SD 0.9), p=0.02. The median total training time for the inexperienced was 48 (range 14 - 63) days, for the intermediate 19 (range 7- 61) days, and for the experienced 25 (range 4 – 60) days.	There was statistically significant difference between the inexperienced and the experienced surgeons in four out of five tasks. The scoring system is easy assessable and can be used for summative and formative feedback.
3	Develop and validate a curriculum for laparoscopic supracervical hysterectomy (LSH).	Single-centre, prospective, cohort study. The curriculum consisted of a theoretical part (written test), a practical part using a VR-simulator and a surgical procedure (LSH), assessed with GOALS and CAT-LSH. Twelve trainees were included (Group 1). The LSH scores were compared with a trainee group without theoretical and practical training (Group 2). The study was performed in March 2015 - Aug 2016.	Ten trainees completed the curriculum. Mean duration of the training period was 57 days (SD 26), and mean training time was 173 min (SD 49). Mean CAT-LSH score was 42.1 (SD 6.9) in Group 1 and 34.8 (SD 4.3) in Group 2, p= 0.009. Mean GOALS score was18.5 (SD 5.8) in Group 1 and 13.6 (SD 3.3) in Group 2, p=0.027.	Trainees who completed the curriculum had a higher performance score on their first hysterectomy, compared to trainees who did not perform structured training.

5. Introduction

The topic of this thesis is assessment of skills in minimally invasive surgery, proficiency levels on a laparoscopic simulator and a curriculum for laparoscopic supracervical hysterectomy.

The introduction starts with a short presentation of the history of surgical education, different surgical training models and assessment tools, learning theory, and legal and ethical issues. The introduction ends with a description of the advantages of proficiency-based training for patients, trainees, and the employers.

5.1 Surgical education

We expect to be treated by competent doctors if we need medical help and surgery. This competence is a combination of knowledge, technical skills, decision making, communication, and leadership skills. Competencies can be learned and feedback is decisive to achieve good learning (1).

The Hippocratic Oath is an ethical code attributed to the ancient Greek physician Hippocrates, and has been adopted as a guide to conduct by the medical profession throughout the ages and still used in the graduation ceremonies of many medical schools. The text of the Hippocratic Oath (approximately year 400 BC) has been reviewed and revised frequently to fit the changes in modern medical practice.

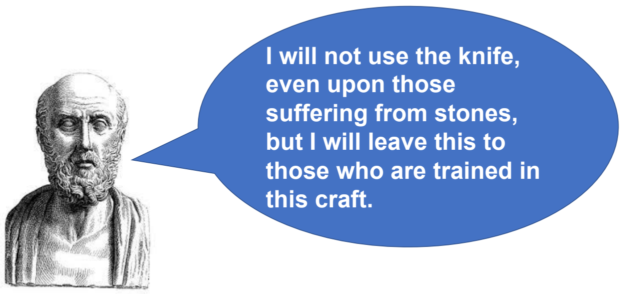


Figure 1. Hippocrates, the father of medicine.

The statement *"I will not use the knife even upon those suffering from stones, but I will leave this to those who are trained in this craft"*, has been interpreted to mean that the physician should yield to "better professionals" whenever needed, and whenever such professionals are available. In the modern context of medical specialization, surgery should be performed by practiced surgeons. Physicians must practice to the extent of their ability but not beyond. "*It is important for us to know our limits and seek help of experts, as needed* " (2).
The first residency programmes for surgical training were introduced in Germany in the late 1880s and adopted in 1889 by William Steward Halsted in the United States (3). Halsted (1852-1922) was one of the pioneers in surgical training. In 1904 he published the paper "The training of the surgeon". This was the first documented surgical training programme. In contrast to the German system, Halsted made the trainees, and not the teacher or professor, the favourite for attention (4). He started safe surgery when he introduced principles, known as Tenets of Halsted; gentle handling of tissue, meticulous haemostasis, preservation of blood supply, strict aseptic technique, minimum tension of tissues, accurate tissue apposition, obliteration of dead space (5, 6). Surgical training has been based on the apprenticeship model for centuries. Skills are transferred from master to apprentice (7). It is based on the assumption that the expert level is reached through experience (8).

5.2 Development of laparoscopy and practical training

Surgery requires incisions into the body enabling examination, correction or removal of organs, structures and tissue. Open surgery is referred to as a technique that requires large incisions, and minimally invasive surgery, a technique that requires small incisions where we can introduce instruments and examine, correct or remove organs, structures and tissue. Laparoscopy (gr: lapro-abdomen, scopein-to examine) is a minimally invasive surgical technique that allows the surgeon to access the abdomen and pelvis. It is also known as keyhole surgery and is carried out under general anaesthesia. Through small incisions in the abdominal wall, carbon dioxide gas is passed into the abdomen to expand the abdominal cavity. A laparoscope, a small tube with a light source and a camera, is then introduced. The images of the abdominal cavity and pelvis is transferred to a 2D monitor in the operating theatre. To be able to introduce working instruments, 2-4 additional incisions of 5-10 mm length are made in the abdominal wall. They are placed in different parts of the abdomen, depending on the planned procedure. At the end of the procedure, the gas is removed, and the small incisions closed. Laparoscopy is beneficial to the patients compared to open surgery in terms of reduced risk of complications such as wound infection, postoperative abdominal wall hernia and venous thrombosis, shorter postoperative convalescence and a better cosmetic result. Following laparoscopic surgery, the patient can often go home the same day. This is not only due to the technique, but also improvement in the pre- and postoperative medication and the anaesthesia methods.

Minimally invasive surgery is possible thanks to the work of pioneers. Technological advances like the development of the laparoscope made laparoscopy feasible, and a turning point in surgical history. The gynaecologist Kurt Semm, (1927-2003) contributed to the development of further necessary technology, such as devices for thermic coagulation and electronic insufflation, and invented procedures for extra- and intra corporal knotting to achieve haemostasis during laparoscopy. Initial laparoscopy was mainly performed as a diagnostic procedure. Since then, the use of laparoscopy in more advances procedures has evolved gradually. In 1981, Kurt Semm performed the first laparoscopic appendectomy and in 1984 the first vaginal hysterectomy with laparoscopic assistance (9, 10). In 1987, the first laparoscopic cholecystectomy was carried out, and in 1988 the first laparoscopic hysterectomy was performed by Harry Reich in Pennsylvania (11). The first laparoscopic hysterectomy in Norway was performed in 1991, by Anton Langebrekke (12).
Skills like manual dexterity, knowledge of anatomy, pathology and surgical techniques are mandatory in open surgery. The implementation of laparoscopic surgery led to a set of new skills to master. Unlike open surgery, the tissue during laparoscopy is handled with long and stiff instruments and the tip of the instruments is not directly visible, but seen on a 2D screen. The laparoscope provides the surgeon with a fantastic vision.

Figure 2. A laparoscopic procedure.
Photograph: Lene Midling-Jensen

Laparoscopic surgery is team work, and navigating the camera by the assistant is essential to achieve this vision. The camera holders thus need training too. Eye-hand coordination due to the fulcrum effect (the tip of the instrument moves in the opposite direction to the surgeon`s hand due to the pivot point) is challenging the tissue manipulation and force transmission.

Following the implementation of laparoscopy, faculty members questioned the training paradigms that had served open surgery well for centuries. In the Halstedian approach, trainees achieved surgical competency by performing a large number of surgical cases. This method did not seem to be suitable for the new type of skills required for minimal invasive surgery (13). Laparoscopy requires an important initial learning process to avoid complications, and initiatives to change and adjust surgical education arose. In 1985, Kurt Semm designed a didactic tool for laparoscopic training. The tool was a box with holes for the instruments, a camera and different objects that could be manipulated by laparoscopic instruments inside the box. The camera projected the inside of the box on a screen. With the help of the screen and laparoscopic grippers and scissors, the laparoscopists in training practised surgical skills outside the operating room.

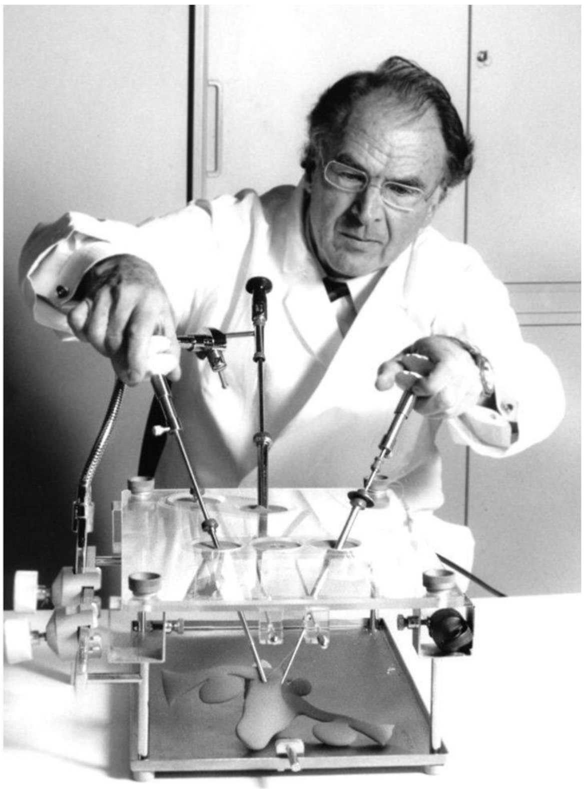


Figure 3. Kurt Semm in 1985 demonstrating the Pelvitrainer,
Reprinted with permission: Department of Obstetrics and Gynecology. University Clinic of Kiel.

Semm published papers, books, films and slides to teach and inform interested colleagues about the technique (10). Laparoscopic surgical training courses where organized in Europe and throughout the USA (10). In the Nordic countries, our Federation of Societies of Obstetrics and Gynaecology, NFOG, decided to have common Nordic standards of clinical practice and assembled a reference group that published clinical guidelines for endoscopic gynaecological surgery (14). Training and education were dedicated one chapter in this report. Different levels of hands-on training on models in pelvic-trainers to practice hand-eye coordination with camera and monitor were suggested (14). This guideline recommended how to train, but not to what level and how to assess, if necessary, what skills were achieved. During the 1990s, technology advanced and it was possible to enhance learning using virtual (VR) reality simulators. A VR simulator is designed as a software program running on a computer which is connected to a user interface. According to the Healthcare Simulation Dictionary, virtual simulation is *"the recreation of reality depicted on a computer screen"* (15). The participant in VR simulation uses instruments with or without haptic feedback, foot pedals, and eventually VR glasses to interact. They can select tasks from a menu in a curriculum. The task can be a surgical procedure or basic skills such as moving different objects around in a particular pattern. When the simulation task is completed, an automated report provides feedback that is objective and standardized using pre-set learning objectives. There are different simulators and on the market, LapSim (Surgical Sience, Gotenborg, Sweden), Lap Mentor (Simbionix ltd, Beit Golan, Israel), MIST-VR (Mentice AB, Goteborg, S), and ProMis (Haptica, Dublin, Ireland). The value of simulation-based training is the ability to practise different parts of a laparoscopic task, and thereby shorten the learning curve. This requires that the skills obtained on the simulator can be transferred to real life surgical procedures. Several research groups have evaluated skills transfer from virtual reality simulators to operations on humans by randomised controlled trials. The intervention groups, the participants with structured training before entering the operation room, perform surgery faster and more safely than participants without training (16-24).

In addition to box trainers and VR-simulators, other training possibilities to develop laparoscopic skills are also available. Cadaver-based instruction is the main instruction tool for anatomy that has been practiced for hundreds of years. Cadaver dissection remains essential in the anatomy curriculum in medical schools, even though modern technology and new teaching methods are available. There is a need for research to evaluate the suitability of new teaching methodologies in new curricula, student perception of integrated and multimodal teaching paradigms, and the ability to satisfy learning outcomes (16). Animal models have a resemblance to human tissue and are typically used on courses in advanced laparoscopic skills. The ability for surgeons to train surgical procedures on cadavers and animals are limited due to ethical reasons, availability, and resources.

Three-dimensional printing allows conversion of digital 3D- models into physical components. It is possible to print a model of the organ we are planning to treat, and that way prepare by performing the procedure on the model of the patients´ organ/tissue before commencing to the procedure on the patient. Although the technique has been implemented for research purposes in some centres, the technology is expensive and not commonly used in surgical training yet (17).

During the past decades, ongoing advances in computer graphics, software, and hardware design have refined medical simulators to offer life-like, or patient-specific simulation, replications of medical and surgical procedures. Such technology requires imaging incorporation of patient specific images, such as computer tomography (CT) and magnetic resonance imaging (MRI) data, into the simulator. It allows preoperative hands-on rehearsal on the upcoming patient. Patient specific VR-rehearsal marks a shift in the use of VR-simulators. The concept of simulated patient-specific rehearsal allows practising a specific event, a planned surgical procedure, opposed to merely act as a generic tool to practice a specific skill (18).

The next logical step in surgical education is to implement the different training possibilities into a context; a curriculum. Some curricula contain a technical skill training programme with basic skills in a low fidelity model, like the Dutch Cobra-alpha course (19). Others contain VR simulator training including basic tasks or basic skills and procedural tasks. A good curriculum contains a theoretical part and a technical part (20, 21). The theoretical part can be self-learning materials like the online program in the General Endoscopic Surgical Educational and Assessment (GESEA) program (22) in the Winners project from the European Society for Gynaecological Endoscopy (23) and Fundamentals of Laparoscopic Surgery (FLS), from the Society of American Gastrointestinal Endoscopic surgeons (24). Some curricula have lectures (25) and others combine self-learning and lectures (26). Deliberate, repetitive practice is essential for performance improvement (8). But the optimal duration of a training session and the optimal interval between them is unknown. The learning principle of massed versus distributed practice has been studied, and there is good evidence that practice interspaced with periods of rest (distributed practice) leads to better acquisition and retention of skills, compared with practice delivered in continuous blocks with little or no rest in between (massed practice) (27-29). Experts have suggested that a one-hour training session might be a good practice (8). Although the evidence for this assumption is scarce, training programmes in a busy clinical setting needs to be practical and feasible to be carried out. Organizing one-hour training sessions once a week is feasible due to the schedule of most departments.

Changes in practice and demographics of both patients and obstetrics and gynaecology have not resulted in substantial changes in resident education (30). Although the effect of surgical training has become known following several studies during the last couple of decades, completing surgical training or curricula is not yet mandatory in the resident training, and training is rarely systematically included in the residents’ and consultants’ schedule.

5.3 Assessment of competence

Education is lifelong. It is however necessary to ensure that certain standards are reached at certain points in a medical doctor's career. Assessment is a central part of our medical education. Decisions and judgements about trainees are often based on clinical practice, number of procedures, rating scales and examinations. The aim is a progression in levels of competency. Trainees move from being inexperienced to become experienced and capable of unsupervised practice. It is necessary to understand the strengths and limitations of the assessment tools utilized to evaluate the competence of a trainee. The number of performed procedures are often used for such assessment, but the number itself is of relatively low value. Assessment during training within a medical field is performed to evaluate increased competence. This is performed by evaluation of whether the trainee is making progress, whether he/she safely can perform a procedure, the readiness to perform key tasks, and what is performed well and what needs to be improved. The tools used to perform such evaluations have become increasingly relevant. At the end of assessments, decisions can be made about the ability of residents to continue with their surgical training, move into the operating room and perform surgery by themselves.

Validation theory has roots in psychology and pedagogy. It is defined as "*the property of being true, correct and in conformity with reality"*, and is a measure that describes whether a test is actually measuring what it claims to measure, and tests the competencies it is designed to test (13, 31). A number of validation benchmarks have been developed, these include construct validity, discriminative validity, concurrent validity, predictive validity, face validity and content validity (13).

Construct validity refers to the extent to which a test actually tests what it is designed to test. *Discriminative validity*, as a subgroup of construct validity, describes if the test is able to discriminate between proficiency levels, thus whether the sensitivity of the test is sufficient. A common example is the ability of a test to differentiate between groups with a different experience, for example experts and novices performing a given task. If the results of the registration of the metrics on the VR simulator show a difference between the groups, we conclude that the assessment has discriminative validity.

Concurrent validity refers to an evaluation of to what extent the relationship between the test score and the score on another instrument purporting to measure the same construct are related. As an example, the concurrent validity is of importance when introducing new assessment tools to replace a pre-existing *"gold standard"* assessment tool.

Predictive validity describes the extent to which a test will predict future performance, hence describing the transferability from one system to another. An assessment tool will have predictive validity if it predicts who will perform actual surgical tasks well and who will not. Predictive validity is most likely to provide a clinically meaningful assessment. The other validities focus on the assessment of the training or test rather than the clinical outcome.

Face validity refers to the extent to which the test actually resembles what happens in a real-life setting. It is typical to ask experts what they think about a simulator or part of it. If they are positive, it is said to have face validity (13).

Content validity is the extent to which the assessment material is meaningful and appropriate as a target for the assessment, like whether the test contains all the steps and skills necessary in a procedure. It is, as face validity, a subjective opinion (13).

Reliability is all aspects of reproducibility or consistency of a test. Consistency is the extent to which a test yields the same results when used under similar conditions or by different examiners (13).

Today many use construct validity as the only form of validity when different training tools are evaluated (32). We need results from a variety of sources to support an evaluation. Five sources, content, response process, internal structure, relationship to other variables and consequences, are noted by the Standards for Educational and Psychological Testing as expedient areas to seek validity evidence (32, 33):

1. Content: Do instrument items completely represent the construct (relevance of the test/simulator with its intended use).
2. Response process: the relationship between the intended construct and the thought processes of subjects or observers.
3. Internal structure: acceptable reliability and factor structure.
4. Relation to other variables: correlation with scores from another instrument assessing the same construct.
5. Consequences: do the scores really make a difference (34)?

Increased attention to the systematic collection of evidence for scores from psychomotor instruments will improve assessment in research, patient care and education (34). The framework of sources can be helpful when an assessment tool is chosen, and for estimations of costs and necessary resources for assessment, feedback and implementation.

Assessments can be designed to be either summative or formative (35).

Summative assessment is designed to evaluate the abilities without necessarily providing feedback for further learning. Exams and certifications are typically summative feedback.

*Formative feedback i*s designed to be a learning opportunity, and highlight the learners' skills, and provide feedback on strengths and weaknesses. It provides information regarding a learning curve and progress towards pre-set benchmarks.

In the second half of the 20th century, the "performance-as-competence" paradigm led to an increased use of a new form of assessment in medical training, known as Objective Structured Clinical Examination (OSCEs) (36). Students were observed at different stations including different tasks like obtaining a medical history, physical examination, or management of different medical conditions. The evaluation was performed using a standardized checklist. Objective Structures Clinical Examinations are one of the most widely used assessment methods in medical education (36).

Surgical skills have often been assessed on-site during surgical procedures. Different assessment tools are developed to make this evaluation as objective and structured as possible. Numerous assessment tools are available for both practical and non-practical skills. Some of these assessment tools are (the list is not complete): "None-Technical skills for Surgeons", "Briefing Intraoperative Teaching Skills for Surgeons", "Briefing Intraoperative Teaching", "Debriefing" and "the Five-Step Feedback Tool For Surgery (37).

The Objective Structured Assessment of Technical skills (OSATS) has been validated and used for open surgery as well as laparoscopic procedures (38). Other validated assessments tools for surgery include Ottawa Surgical Competency Operating Room Evaluation (O-SCORE), the Operative Rating System, or the Zwisch scale, Global Operative Assessment in Laparoscopic Surgery, GOALS, Competence Assessment Tool (CAT), Vaginal Surgical Skills Index (47-50).

Diagnostic reasoning is an essential part of clinical competency. Feedback works at its best when it inspires reflection on behaviour and performance (39). The ideal assessment tool produces reliable, valid results and is furthermore practical (40). In the apprenticeship model the assessment was the responsibility of the trainers and was largely subjective (41, 42). Structured assessment can provide trainees with useful knowledge about their performance, and feedback can shorten the learning curve (43).

Technical skills can be defined as a set of abilities or knowledge used to perform a practical task. In most cases, the acquisition of advanced technical skills requires specialized training and education, which require both time and other resources such as training facilities and systems. The certification required to become a specialist is changing and include more and more simulation training (44). At the end of any teaching intervention, it is essential to assess what learners can do. The essential measure is the output. The input, such as costs, time, and necessary resources to implement the training, is also interesting in order to plan and implement efficient training systems or curricula.

5.4 Miller’s pyramid of assessment

In 1990, the psychologist George Miller described a method for ranking clinical competence in the workplace in an educational setting (45). He argued that the traditional assessment of medical students relied too much on testing their knowledge, and not how they would behave in real-life consultations. Miller described a pyramid with four levels; knows, knows how, shows how, and does (45). To truly know whether our trainees are achieving what we want them to achieve, we must evaluate them in the setting we expect them to deliver the competence.

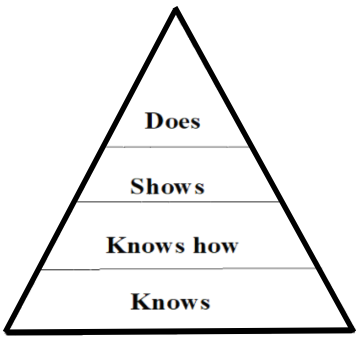


Figure 4. Miller`s pyramid of assessment.
Illustration: Jeanne Mette Goderstad

With this framework, it is possible to identify learnings objectives, link them to the different steps of the pyramid, choose a suitable assessment tool, and perform testing. Every skill begins with a knowledge component essential for being a professional. On the lowest level of the pyramid is “knowledge”. This level can be organized as knowledge courses and tested by written exams and multiple-choice questions.
In the “Knows how-level”, the ability to apply knowledge is achieved. This level can easily be assessed by essays, theoretical clinical problem-solving exercises and extended multiple choice questions.

The third level, "shows how", represents clinical skills competency. At this stage, the knowledge is used to perform specific actions. This level can be evaluated using simulations, standardized patient exercises and practical task in models. The performance can be evaluated with assessment tools. Finally, on top, "The Does level", is reached when the clinical performance is performed. The highest level of the pyramid can be evaluated more or less objectively by direct observation, simulated patients, video recordings that are assessed by checklists, assessment tools like OSCEs, OSATS and GOALS, portfolios and audits. To reach this level within laparoscopic surgery this requires both knowledge and practical skills. Proficiency based education is developed by experts performing a defined target procedure or a set of surgical skills. Their performance is the goal for the training, and can represent the summative feedback. The trainees are certified when they reach the pre-set proficiency level.

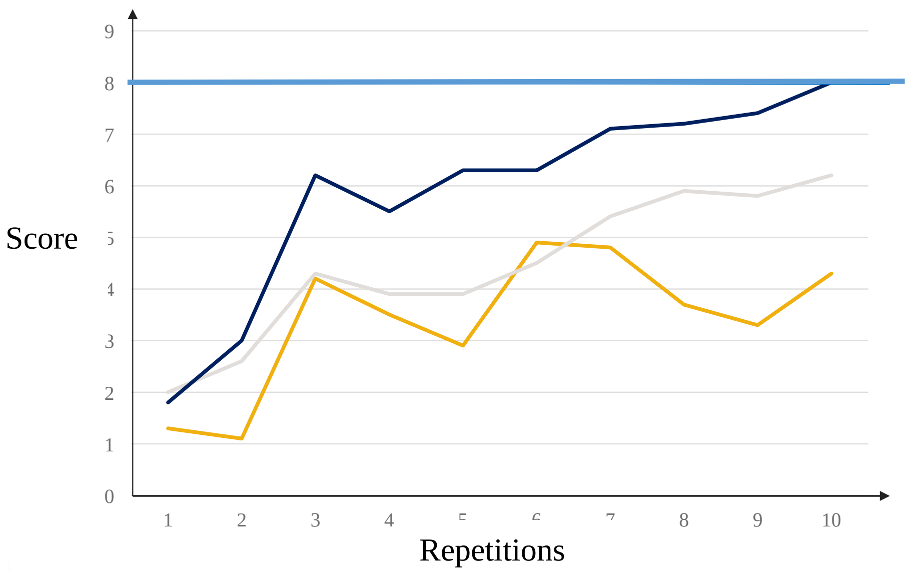


Figure 5. The proficiency level – "good to go".

The blue line illustrates the pre-set proficiency level. Every result below the blue line can be described as "not yet" with is a positive designation reflecting on that the trainees are on their way to become proficient.

It will differ how many repetitions the trainees need to achieve proficiency. Number of repetitions is of less importance. The crucial point is that they reach the proficiency level.

Illustration: Jeanne Mette Goderstad

5.5 Legal and ethical issues.

National and international laws, standards and guidelines for good medical practice regulate medical practice. In addition, patients' satisfaction rates and patient associations might affect the medical practice. Quality control has generated considerable interest from health authorities, health personnel and patients themselves in recent decades (46) (47).
Several National Quality Registers are used in Norway, in order to follow up the effect and risk of complications of medical and surgical treatment. Hence, both national, local and personal quality data are available for treatment of different conditions. From the Norwegian Gynecological Endoscopy Register, national and hospital overview procedures and complications can be obtained. Doctors must be given the possibility to perform a good job when it comes to working conditions, like enough time to carry out a professional job for patients, with procedures and knowledge. The doctors must make themselves available for evaluation. The trust and confidence that doctors contribute to good health care are essential to the public, the doctors and the doctors employers (48). There are different possibilities to assess doctors´ competence. Monitoring surgical performance with Cumulative Summation Techniques, CUSUM, is one example, which is recommended for use as a personal audit at an individual level (49).

In 1998 Professor Trond Buanes published an article in a Norwegian newspaper reminding employers at hospitals and doctors themselves that they have a responsibility to implement structured training of surgical skills. He further claimed that surgeons without the necessary training should not perform complicated operations (50). Consequently, the surgeon has a personal responsibility to possess the necessary competence to perform a procedure. But, how does a surgeon know the answer of the following question: *"When are you good enough for your patients?"* (51). Surgeons who perform procedures without additional training are three times more likely to have at least one complication compared with surgeons who attend additional training (52). Furthermore, a certain number of cases are mandatory before surgeons attain proficiency. It is predicted that 90% of the injuries occur during a surgeon`s first 30 cases (53).

Assessment and certification of skills within minimal invasive surgery are considered a gold standard for assuring that a surgeon has acquired and retained a certain level of knowledge and skills. Hospitals training gynaecologists rarely use available educational programmes for endoscopic surgery. It is not considered a learning priority element for the residency programme (54). Emphasis needs to be placed upon systems, in addition to the performance of individuals. Without training, the hospitals' routines and the doctors' position are poor when it comes to complaints from patients and supervisory authorities. With training, the position might have a favourable perspective since training programmes are expected to improve patient safety and outcome. Moreover, the medico-legal consequences can be affected when a significant number of surgeons possess different diplomas like GESEA and ECRES (54).

It has been assumed that doctors would remain competent throughout their professional careers by taking postgraduate courses and work in relevant practice. The requirements from patients, society and peers make a professional career and its assessment much more complex today, than it was in the past. In many countries physicians must demonstrate their engagement in lifelong learning by choosing and participate in continuing medical education (CME) and in lifelong learning (55). New credit systems are needed to measure a CME activity by its value, aiming to improve the physician`s knowledge base, competence and performance in practice.

5.6 Advantages for patients, trainees, and employers.

The aim of surgical training is to reduce the risk of surgical injuries and complications and increase patient safety. When proficiency- based surgical education is implemented in practice, a patient will experience treatment from a surgeon with skills that are assessed and evaluated good enough for the surgical procedure she is about to have. There will always be differences between each individual health worker, but the performance of the surgery will be conducted to a standard evaluated as good enough, and the surgeons are supervised, tested and "signed in" to perform surgery. If the planned procedure is postponed, the same standard will be applied in case the procedure is performed by another surgeon on another day or at another time of the day.

In the book "Cutting edges in surgical training" Shekhar Kumta stated that unsupervised or poorly supervised trainees may adopt ineffective and risky strategies dealing with surgical problems beyond their capabilities (56). This risk is reduced when surgeons are signed in to surgery, by having a certification of a set of skills and required defined proficiency levels before they enter the operation room.

Proficiency based surgical education with simulation is designed for the benefit of the learner. In skills laboratory settings, the trainee can have deliberate practice, mistakes are allowed, the environment is safe, they have 24-hours availability, variation of cases and exercises, a possibility for peer teaching, and assessment is standardized.

Trainees find simulation useful. In a Dutch study, 75% of the residents found endoscopic skills training outside the operation room useful (57, 58). It causes transparency between residents when criterion-based goals and requirements for trainees to complete modules to a proficiency level before performing any procedure in the operating room are established (59). When the training is completed, the trainees have documented skills and competence and are prepared for the procedure. Trainees may therefore encounter the surgery as less stressful and hence make better decisions during the surgical procedure.

Previous studies have shown that surgeons perform safer and faster surgery following the implementation of proficiency-based education (60). Since they are more effective and safer in the operating room it, can lead to financial benefits for the hospital. Many trainees do not practice enhancing the surgical competence by themselves, so the employers should establish a mandatory curriculum (68,70,71). Stafanidis et al. registered that the attendance in skill training rose from 6% to 71% when time was dedicated to training and supervising personnel were available (61). The employers must further plan for training in the work schedule, make training facilities available, estimate the costs and prioritize surgical training in their budgets in line with other costs.

6. Objectives of the thesis

The main objective of the studies was to develop a validated training programme in laparoscopic surgery feasible in the education of trainees in gynaecology and obstetrics in Norway. The programme should be proficiency-based and “sign the participants in” to the operating theatre when they had reached the pre-set proficiency level for the practical training. Objective feedback should continue in the operating room by use of validated assessment tools. In order to put together this training programme, we aimed to:

- develop and validate assessment tools for the surgical performance
- define pre-set proficiency levels for the practical training
- develop a rating scale
- put together a training programme including a theoretical part as well as practical tasks
- validate the training programme by evaluation of the surgical performance of participants who completed the curriculum, compared to performance of participants without systematic preoperatively training.

7. Material and methods

7. 1 Formalities and approvals

Prior to the start-up, the studies were approved by the Head of The Department of Gynaecology at Oslo University Hospital, and the Head of the Research Centre for Obstetrics and Gynaecology at Oslo University Hospital. The aim of the studies was to register and evaluate aspects of surgical training that are considered an integrated part of the healthcare service. According to The Research Handbook 2021, chapter 6.4; The demarcation between research projects that need approval from Regional Committees for Medical and Health Research Ethics, REK, quality studies and other research projects, studies that are subject to the overall health legislation, do not need approval from REK (62).

Prior to inclusion, all study participants received written information about the study signed an informed consent form for study participation.

The patients received preoperatively written and oral information about the study and they signed an informed consent form approving that their surgical procedure was video recorded and evaluated afterwards.

7.2 Procedure selection

In general, a similar curriculum as ours, using pre-set proficiency levels based on the experienced surgeon`s performance, could be developed for any laparoscopic procedure. We chose hysterectomy, both because it is a common procedure, and because the study participants were at a stage in their surgical education where performing hysterectomy would be the next step. By choosing hysterectomy, we consequently had access to a large number of procedures and ensured that the participants were motivated for the training. Hysterectomy is an essential operation for gynaecological trainees to learn and master, and it is also on their procedure list for becoming specialists. Laparoscopic supracervical hysterectomy (LSH) is an option for women suffering from benign gynaecological conditions. When there is no need to remove the cervix or removal is not desired, it is a less invasive procedure compared to total laparoscopic hysterectomy. The procedure is less complicated than the total laparoscopic hysterectomy, which makes it a suitable hysterectomy option for the trainees. The laparoscopic supracervical hysterectomy is an alternative to total laparoscopic hysterectomy and it is therefore necessary for gynaecologists to know both procedures and departments to offer women both methods.

The Department of Gynaecology at Oslo University Hospital is the largest gynaecological department in Norway, both when it comes to annual number of hysterectomies and also number of trainees (34 trainees). Furthermore, supracervical hysterectomy was considered to be the most suitable operation for trainees who were about to learn hysterectomies at Oslo University Hospital, Ullevål.

7.3 Assessment tools selection

To evaluate if something is adequate, we need to have an opinion/definition about what is good enough, and furthermore be able to register and assess what we are evaluating in an objective way. In order to evaluate the surgical procedures in our studies, we used a procedure-specific rating scale and a general assessment tool for laparoscopy.

There was no available procedure-specific rating scale for laparoscopic supracervical hysterectomy at the time we planned to conduct the study. We assumed that a procedure-specific rating scale might add useful information and facilitate summative and formative feedback. In order to develop Competence Assessment Tool – Laparoscopic Supracervical Hysterectomy (CAT-LSH), we needed a standardised strategy for hysterectomies suitable for trainees. To develop this strategy, we interviewed three recognised international experts with an interest in education and recognition in endoscopic surgery from the European Society of Gynaecological Endoscopy (ESGE) and the American Association of Gynaecological Laparoscopists (AAGL). We were inspired by Competence Assessment Tools for different procedures made by colleagues from Imperial College in London, and we designed CAT-LSH in the same way (Appendices) (63). The minimum score for each step is 4 and the maximum 16, giving a minimum score of 16 and a maximum of 64.

Global Operative Assessment of Laparoscopic Skills (GOALS), is a general assessment tool with five general domains (depth perception, bimanual dexterity, efficiency, tissue handling, and level of difficulty) and has previously been validated for cholecystectomy, ventral hernia repair and appendectomy (8,11,14). By adding the item and degree of difficulty, we had the possibility to equalize the procedures by correction of case mix (11,14). The minimum score is 6 and the maximum score is 30. The possibility to evaluate degree of difficulty and the straightforward structure of the assessment sheet were among the reasons we chose GOALS as the general assessment tool (Appendices).

7.4 Task selection

At Oslo University Hospital, we had the LAPmentor Express, Simbionix, 3D Systems available for conducting the studies. The simulator had basic and procedural tasks, including a hysterectomy. It was a portable, 2D non-haptic feedback simulator with a manageable size which made it possible to transport the simulator in a car in order to travel between the hospitals to conduct the study participants’ training sessions.

Figure 6. The training setting.

Photograph: Lene Midling-Jensen

We wanted the participants to practice skills that facilitated the performance of the different steps of the hysterectomy, like use of both hands, accuracy, depth perception, instrument manipulation, tempo, eye-hand coordination, tissue-handling, coagulation, cutting and to follow a strategy for the whole procedure. We looked through the different tasks on the simulator and ended up including the following in the training programme:

Task 1: Two-handed manoeuver

Task 2: Peg transfer

Task 3: Pattern cutting

Task 4: Salpingectomy

Task 5: Hysterectomy

The hysterectomy was modified to imitate a supracervical hysterectomy. Following the initial steps of a hysterectomy, the participants marked with a hook the the correct level of amputation of the cervix on the anterior and posterior part of the exposed cervix.

Information about the 5 tasks follows:

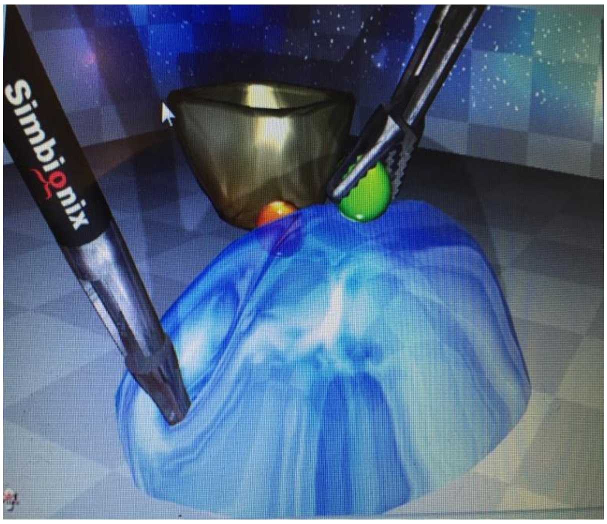


Figure 7. Task 1: Two-handed manoeuver

Photograph: Jeanne Mette Goderstad

This is a coordination task involving speed and precision. The objectives were to improve advanced bimanual skills, to practice instrument manipulation, practice eye-hand coordination and acquire tissue-handling skills.

In this task nine balls were imbedded in jelly. The task included retraction of jelly to expose all balls. A correctly exposed ball changed the color from red to green. All green balls had to be grabbed with the other instrument and placed into a bag. The parameters measured were time (s), number of balls in the basket (n), total path length (cm) and instrument movement (number). In addition, number of errors wereregistered (only green balls should be grabbed).

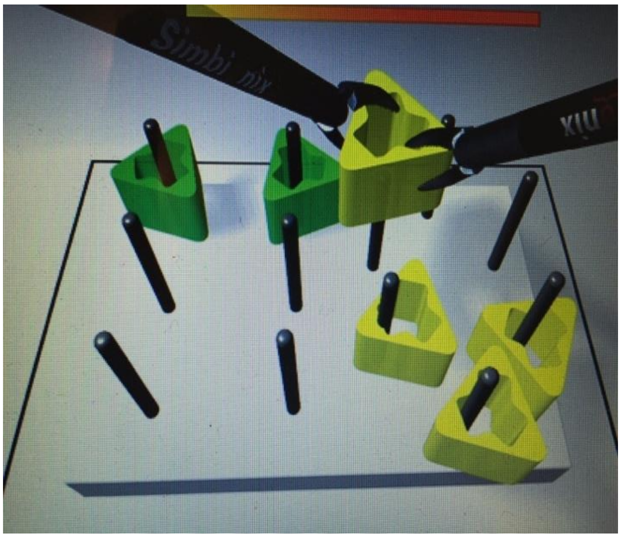


Figure 8. Task 2: Peg transfer
Photograph: Jeanne Mette Goderstad

The participants lifted six objects from a pegboard with their left hand, transferred the object to their right hand, and placed them over the pegs on the right side of the pegboard. The process was then reversed. The objectives were improved eye-hand coordination, use of both hands and improved depth perception. The parameter measured in this task were total time (s) and number of successfully moved objects (without loss and correctly placed on the pegboard) (n).

Figure 9. Task 3: Pattern cutting.
Photograph: Jeanne Mette Goderstad.

The objectives of this task were use of both hands and accuracy. The participants used a grasper to apply traction exposing the best angle for the dominant hand to cut the marked circle with accuracy

The parameters measured were total time (s) and errors (any deviation from the drawn line).

Task 4: Left side salpingectomy

The objectives of this task were to coordinate the coagulation and the cutting.

The participants used a grasper, scissors, and bipolar forceps to remove the left tube. The total time used on the task (number of minutes) was registered. In case of an error (bleeding), it had to be corrected before commencing the salpingectomy.

Figure 10. Task 5. Modified laparoscopic supracervical hysterectomy.
Photograph: Jeanne Mette Goderstad.

The participants were introduced to a step-by-step strategy of the procedure. It started on the left side and included (64)

1. Identification and division of the round ligament
2. Identification of the anterior leaf of the broad ligament and progressive cauterization of the ligament towards the middle medially paying attention to the bladder
3. Coagulation and division of the proper ovarian ligament and the fallopian tube
4. Division of the posterior leaf of the broad ligament
5. Identification, coagulation, and division of the uterine vessels
6. Step 1-5 was then performed at the right side
7. The cervix was exposed and the participant marked the correct level of amputation

In this task, total procedural time (min), total path length (cm), instrument movements (n) and errors (bleeding and improper respect of tissue/tissue handling) were registered. The registration started when the participant took hold of the left round ligament and ended when they had marked the amputation level of the cervix.

7.5 Rating scale

We used the results of the parameters of the four last repetitions of each task performed by the experienced participants on the simulator as the base for the rating scale

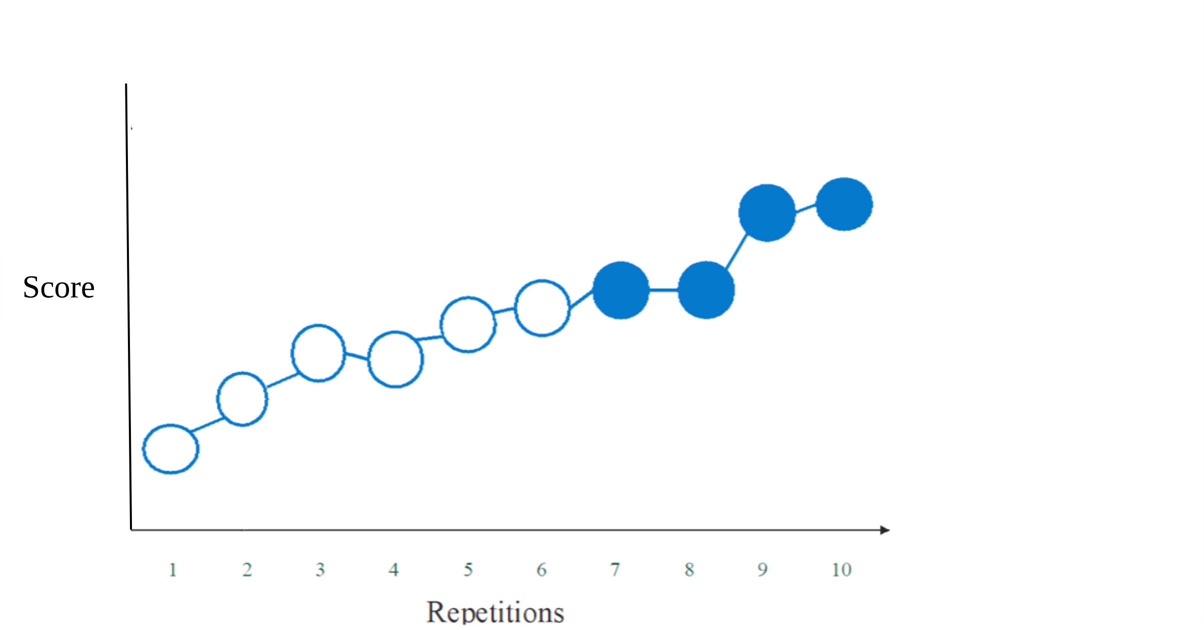


Figure 11. The base for the rating scale. The experienced participant's score.

Illustration: Jeanne Mette Goderstad

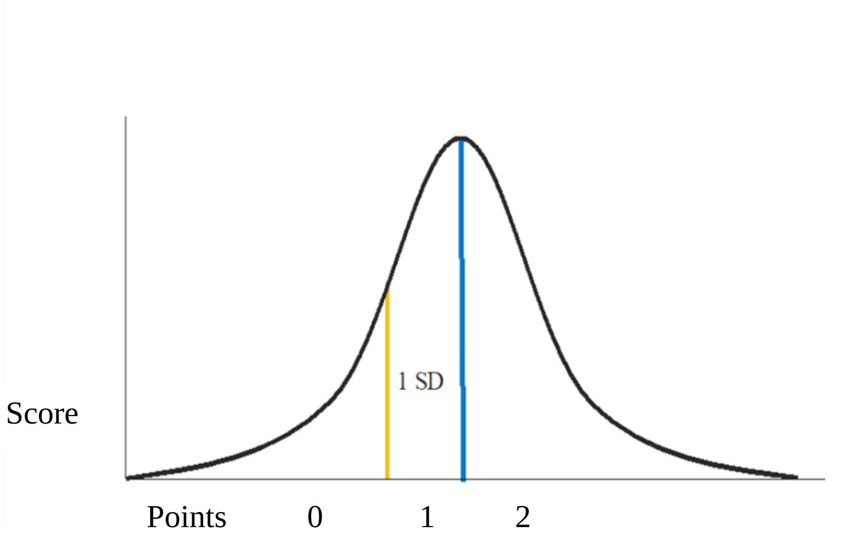


Figure 12. The rating scale calculation. The mean and SD were registered for each parameter in each task. Performance equal to, and higher than the mean score of the experienced surgeons in each parameter was defined as two points. One SD decrease from the mean resulted in a score of one point. Every score below one SD gave 0 point.

Illustration: Jeanne Mette Goderstad

The scores from each parameter in the tasks on the simulator were added to give the total task score. Since the different tasks had different numbers of parameters used for evaluation, the maximum score differed between the different tasks. It was 10 in task 1, 4 in task 2 and 3, and 8 in task 5. In Task 4, we used time (min), since only one parameter was measured.

7.6 Participant selection

We used the following categories:

1. Inexperienced (had performed less than 50 laparoscopic procedures in total, and previously never performed a laparoscopic hysterectomy). These trainees were about to move on to more complex procedures, like a laparoscopic supracervical hysterectomy, and had thereby basic skills.
2. Intermediate experienced (had previously performed more than 50 laparoscopic procedures in total, including more than five laparoscopic supracervical hysterectomies, but not performed total laparoscopic hysterectomies).
3.Experienced (senior consultants performing total laparoscopic hysterectomy and surgery for deep infiltrating endometriosis).

This categorization was based on a previous study on laparoscopic training (65).

The participants were recruited from Oslo University Hospital. When developing the scoring system for laparoscopic skills on the simulator, we added trainees and gynaecologists from the gynaecological departments of Akershus University Hospital and Vestre Viken Hospital Trust, Drammen, to obtain the necessary number of participants. Akershus University Hospital is a public university hospital in the Greater Oslo Region. Vestre Viken Hospital Trust, Drammen is located 40 kilometres outside of Oslo. These hospitals were chosen for practical reasons.

All available trainees at the department at Oslo University Hospital, were informed about the study and invited to participate. We included all available trainees that were eligible according to the inclusion criteria, and planned to continue their employment at the department in the hospital the following six months. Every participant meeting the criteria of inclusion was selected until the required sample size was achieved. The gynaecologists were recruited from the gynaecological department. The inclusion at Akershus University Hospital and Vestre Viken Hospital Trust, Drammen, were performed in a similar way.

7.7 Validation of the assessment tools

We evaluated GOALS and CAT-LSH by using them to assess surgical performance of a Laparoscopic supracervical hysterectomy by unexperienced, intermediate experienced and experienced gynaecologist. The operating assistants were informed about the surgical strategy the participants were supposed to follow, the contents of the two assessment tools and how to fill in the scores. During surgery the operating assistants were supposed to correct the participants if they were about to perform a risky manoeuvre. Otherwise, the participants should complete the operation without supervision. Immediately after the surgical procedure, the assistant filled out the two assessment forms.

All the laparoscopic supracervical procedures were video recorded and stored on separate memory sticks. The recording started when the participant took hold of the left round ligament, paused when the assistant operated on the right side, and started again when preparing for the amputation of the cervix and ended after the amputation.

Two gynaecologists who were highly experienced in laparoscopic surgery performed the video evaluation. Assessment of the laparoscopic supracervical hysterectomy performance was done using CAT-LSH and GOALS. We gave them information about the assessment tools, how to mark the scores on the assessment form, and that they were allowed to watch the videos for as many times as they wanted. They were informed about the standardised strategy of the operation given to the participants, and that the participant were allowed to use grasper, bipolar and scissors. The two experts received envelopes marked with a number, containing the memory stick and the two rating scales. The envelopes were given in a random order, and they received 5-8 at the same time. The experts were blinded to the identity of the participants.

When we had the GOALS and CAT-LSH results, we were interested in the agreement between the operating assistant and the two blinded observers. It was also of interest to know the agreement between the two blinded observers. To calculate these values, we used the interclass correlation coefficient.

7.8 The training sessions

We planned for not more than 3-4 repetitions at each training session on the simulator due to the principle of distributed training. The training was planned to be carried out in between the daily tasks, before and after on call duty, which led to the practicality of short training sessions. We assumed that 3-4 repetitions were practical according to duration of time to complete the session without interruptions.

7.9 Theoretical part of the curriculum

A Danish colleague, Jeanette Strandbygaar, has published a multiple-choice test within basic laparoscopy (Appendices) (66). We contacted her and were allowed to use her test in the curriculum. Since this was theory in basic skills, we added six questions focusing on the laparoscopic supracervical hysterectomy (Appendices). These questions were developed in cooperation with three experienced gynaecologists at the Department of Gynaecology at Oslo University Hospital.

7.10 Methods in the three papers

Study 1: Development and Validation of Rating Scales Used for Laparoscopic Supracervical Hysterectomy.

When we had the two rating scales, CAT-LSH, and GOALS, a pilot study with ten participants was conducted. We ended up including 37 procedures performed by specialists and specialist trainees with different proficiency levels. The participants were instructed to follow the surgical strategy: ligament mobilization, release of the adnexa from the uterus, division of the uterine vessels and the uterus amputation. To standardize the procedure all operations were performed with bipolar desiccation, grasper, and scissors. The procedures were video recorded and stored on memory sticks. Assessment with CAT-LSH and GOALS was made by the assisting surgeon immediately after surgery. They received information about the study and the rating scale before the surgery. The operating assistant who was scheduled to the operating room the day of the procedure was chosen as the operating assistant and assessor. We did not make changes to the operating programme concerning selection of patients. The operation was also assessed by two experts in laparoscopy receiving the memory sticks and blinded to the identity of the participants and their proficiency level.

We compared the mean scores of the different proficiency groups and estimated the interrater reliability between the blinded observers and between the assistant surgeon and the observers.

Study II: **Development and validation of a general and easy assessable scoring system for laparoscopic skills using a virtual reality simulator.**

Study II is a prospective, longitudinal cohort study.

We recruited consecutively, until ten participants had been included in each study group. All participants were given an individual hands-on introduction to the simulator, an oral presentation as well as a video presentation of the tasks. The program consisted of three basic skill tasks (Task 1, 2 and 3), a salpingectomy (Task 4) and a modified laparoscopic supracervical hysterectomy (Task 5). We chose these tasks because together they give the trainee the skills necessary for the steps of the procedure: the ligament mobilization, release of the adnexa from the uterus, division of the uterine vessels and the uterus amputation. The participants performed all tasks during each training session in a systematic order, starting at Task 1 and continuing to Task 2, 3, 4, and 5 consecutively. This was repeated, dependent of available training time, up to maximum of four repetitions during one training session. The training was completed when all the tasks had been performed ten times. The total training period was aimed to last between two and six weeks.

The main study investigator (JMG) was present during all training sessions in order to assist the study participants in case they needed guidance on the simulator system or information regarding the different tasks

Study III: Development and validation of a curriculum for laparoscopic supracervical hysterectomy.

When we had the theoretical knowledge questions, the scoring system, the proficiency levels for training on the simulator, and the assessment tools for laparoscopic supracervical hysterectomy we designed a curriculum. The curriculum consisted of three steps.

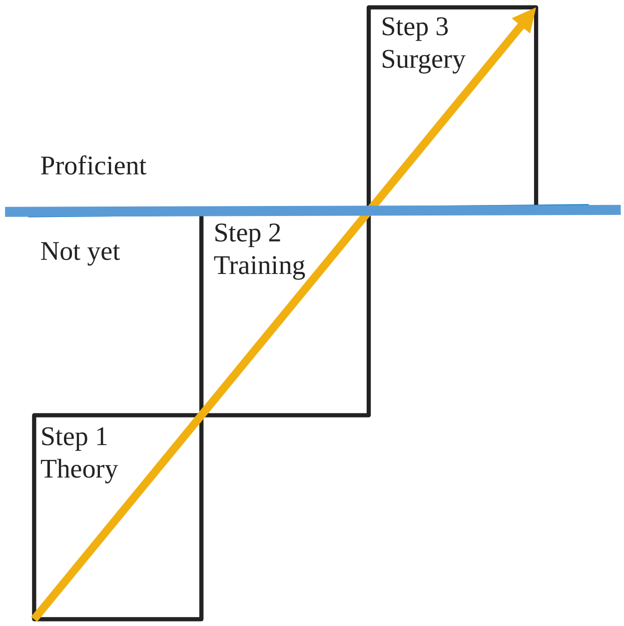


Figure 13. The three steps in the curriculum.
Illustration: Jeanne Mette Goderstad

Step 1, the theoretical knowledge part with a multiple-choice test within basic laparoscopy (66) and six questions related to the laparoscopic supracervical hysterectomy procedure. It was completed as a home exam. Wrong answers were discussed to enable the trainee to give the correct answer.
Step 2, was an individual laparoscopic training program on the the Simbionix LAPmentor Express 3D, VR simulator. The practical part consisted of three basic skill tasks, a salpingectomy and LSH. The participants rehearsed until they reached the pre-set proficiency level based on the scoring system we had developed. We planned for not more than three repetitions at each training session due to the principle of distributed training.
When the participants reached the pre-set proficiency level, they performed a laparoscopic supracervical hysterectomy. (Step 3). We scheduled that the participant operated the next patient planned/set up for a laparoscopic supracervical hysterectomy at Oslo University Hospital, Ullevål. The participants were instructed to follow the surgical strategy: ligament mobilization, release of the adnexa from the uterus, division of the uterine vessels and the uterus amputation. To standardize the procedure all operations were performed with bipolar desiccation, grasper, and scissors. The procedures were video recorded and stored on memory sticks. The video recording started when the participant took hold of the round ligament on the left side, stopped when the assistant operated on the right side and started again when the amputation was performed. Assessment with CAT-LSH and GOALS was made by two

experts in laparoscopy from the memory sticks. They were blinded to the identity of the participants. The results were compared with the findings from the assessment of video recordings of 11 laparoscopic supracervical hysterectomy procedures, performed by the unexperienced participants in paper I. They had the same level of competence, regarding surgical procedures as the participants in paper III, but not any structured training before the operation.

7.11 Statistics

7.11.1. Test power

Paper I: The sample size calculation was based on a pilot study. We included ten doctors with different surgical experience, who performed an LSH. The procedure was video recorded, assessed by the operating assistant (trainee or consultant) and two blinded experts using GOALS, and CAT-LSH. The categorization was the same as in the main study. We had to include 37 laparoscopic supracervical hysterectomy procedures to achieve a study power of 80% at a level of significance of 0.05.

Paper II: The sample size calculation was based on the variable, "duration of task" from a procedural task, salpingectomy, in a study by Larsen CR et al. (9). The standard deviation of time in Larsen's study is 90 seconds in the inexperienced group, and 40 seconds in the expert group. We assumed that similar standard deviations would be observed in our study. We furthermore assumed that the mean difference in time between Group 1 and Group 2 would be at least 90 seconds. It may be shown that if the true mean difference in time between these two study groups is at least 90 seconds, in a study with 80% test power and a significance level of 0.05, at least 10 physicians had to be included in each group. We consequently decided to include 10 study participants in each study group.

Paper III: The sample size calculation was based on the total GOALS score from a previous simulation study with inexperienced and experienced surgeons. The mean difference in GOALS between the groups was 5.8, and the standard deviation in each group was 4.0. (ref Larsen 12) We assumed that we would obtain the same difference in GOALS score, between the groups in study III and with 5 % significance level and 80% test power at least 10 trainees had to be included in each group.

7.11.2. Statistical analysis

All statistical tests were 2-sided, and $p \leq 0.05$ was considered statistically significant.

Statistical analyses were performed using commercially available software (SPSS version 17.0; SPSS Inc., Chicago, IL).
In study 1, continuous data from two study groups were compared using a 2-sided Independent Samples t-test. We used the interclass correlation coefficient to measure the agreement between the assessors.
To evaluate if the data was close enough to normal distribution, we used the findings in Morten W. Fagerland's paper. (67)

7.12. Ethical considerations

The studies were conducted in accordance with the Declaration of Helsinki: The well-being of the human subject should take precedence over the interest of science and society, consent should be in writing and the participants benefit from research (68).

8. Summary of results

8.1 Validation of the assessment tools GOALS and CAT-LSH (paper I)

We ended up with 37 laparoscopic supracervical hysterectomy procedures being assessed by the assistant surgeon and the blinded observers using GOALS and CAT-LSH. Eleven procedures were performed by inexperienced trainees, 12 by intermediate experienced gynaecologists, and 14 by experienced gynaecologists. The indication for surgery was uterine fibroids and/or menorrhagia, and the median uterine corpus weight was 261g, (range 40 - 820 g).
The assistant surgeons evaluated the intermediate experienced gynaecologists with a higher score than the inexperienced gynaecologists. The differences were statistically significant. This was the case for the assessment with both GOALS and CAT-LSH. The blinded observers came to the same results, the intermediate experienced gynaecologists had a statistically significant higher score than the inexperienced gynaecologists.
The assistant surgeons evaluated the experienced gynaecologists with a higher score than the intermediate experienced gynaecologist. The differences were statistically significant for assessment with both GOALS and CAT-LSH. The blinded observers obtained the same result, a statistically significant difference between the groups, giving the experienced gynaecologists a higher score than the intermediate experienced participants.

The assistant surgeons gave the participants; the inexperienced gynaecologists, the intermediate experienced gynaecologists and the experienced gynaecologists, higher scores as groups, compared to the blinded observers. This difference resulted in a good interrater reliability score, with an interclass correlation coefficient (0-1) of 0.71 for GOALS and 0.75 for CAT-LSH, respectively.
The interrater reliability scores of the two blinded observers were 0.74 for GOALS and 0.85 for CAT-LSH.

8.2 Development and validation of the scoring system (paper II)

The time spend on the simulator to complete the training (10 repetitions of the 5 tasks) differed between the groups. The intermediate experienced group had the shortest median training period, 19 days (range 7-61 days), the experienced group had a median training period of 25 days (range 4 -60), and the median total training time for the inexperienced group was 48 days (range 14 - 63 days),

There was no statistically significant difference between the total score for the inexperienced and the intermediate experienced group in any of the five tasks.
Task 1: Two-handed maneuver
Task 2: Peg transfer
Task 3: Pattern cutting
Task 4: Salpingectomy
Task 5: Hysterectomy

There was a statistical significance in total score between the inexperienced and the experienced group in tasks 1, 3, 4 and 5. There was no statistically significant difference in task 2.

There was a statistical significance in total score between the inexperienced and the experienced group in tasks 1 and 3. No difference in tasks 2, 4 and 5.

8.3 Development and validation of the curriculum (paper III)

We included 12 participants and all of them completed the three steps of the curriculum. The mean duration of the training period (step 1 and 2) was 57.0 days (SD 26.0). The participants spent a mean of 173.0 min (SD 49.0) on the simulator to reach the pre-set proficiency level for all tasks.

The participants spent different times on the different tasks. The mean time on task 1 was 12.5 min (SD 3.7), task 2, 17.8 min (SD 7.7), task 3, 22.7 min (SD 13.5), task 4, 22.9 min (SD 10.1) and task 5, 97.9 min (SD 36.9).

There was also a difference in number of repetitions needed to reach the pre-set proficiency level. The mean number of repetitions for task 1 was 8.8 (SD 3.0), task 2, 7.8 (SD 3.0), task 3, 9.6 (SD 4.8), task 4, 6.5 (SD 2.7) and task 5, 9.1 (SD 3.0) respectively.

Unfortunately, the recording of the procedure failed in two cases.

The inexperienced participants who completed the curriculum had a statistically significantly better score on the video evaluation of the performance of the laparoscopic supracervical hysterectomy, than the inexperienced group without the structured training that we presented in paper 1.

8.4 Overall results

The curriculum we designed is feasible for the education of trainees in gynaecology and obstetrics in Norway.

- the mean performance score of the video-recorded LSH procedure was significantly better for the trainees who had structured training compared to the trainees without.
- the mean duration of the training period was 57.0 days (SD 26.0)

9. Discussion

The studies presented in this thesis focussed on assessment of skills, development of proficiency levels and a curriculum for a gynaecological procedure, viz. the LSH. Some results are specific for training in gynaecology, but other aspects can be transferable for training, rating scales and assessment of surgical skills in general.

Objective assessment tools for surgery have been available for more than 20 years. Simulation based training for minimal invasive surgery has also been available for more than 20 years. However, in our opinion, systematic use of objective assessment tools and surgical simulators is still rarely used in daily practice.

9.1 Methodological considerations

Study participants

In the three studies we included participants that were actively participating in clinical work. They were therefore representative of the current clinical setting concerning who we regard as experts, intermediate experienced and inexperienced gynaecologists and trainees. In study II we registered age, dominant hand and whether the study participants had any previous experience with the LapMentor simulator. This information is mentioned in the publication. When we planned the study, we also wanted to register any sport activities, playing an instrument and gaming among other variables. We thought it might be interesting to register if there was a difference in learning curves and eventually adjust for a set of skills in advance of the training and calculate any training effect on the simulator.

During the work with the studies the “sign them in” principle became increasingly interesting. The participants had different personal characteristics and experience when they started the training. However, they would end up more equal when it came to skills, defined by the pre-set proficiency level, when they completed the training and were signed in to perform surgery. It was of little interest if the registrars were right- or left-handed, men or women, whether they had used the simulator system before or not, were young or older, as long as they had reached the pre-set proficiency level and completed the curriculum. They were “signed in” for surgery and had achieved the competence of skills defined to the procedure they were going to perform. The consequence was that we could have dropped the registration of such variables in our study, and it can consequently be avoided in future studies. The trainees had a different base and a different learning curve. In our training design, we do not need to pay attention to such differences, as long as we have a pre-set training goal and certification with objective assessment tools. Some are fast learners, and some are slow, resulting in different times spent to reach the proficiency level.

We recruited the study participants from our own department in study I and III. All trainees who matched the inclusion criteria and planned to work at the hospital in the next six months were invited to participate. No one declined the offer of inclusion. In this way we obtained a representative population of trainees, given that we recruited from our own hospital. In our opinion, the fact that no one turned down study participation, most likely shows that they are interested in feedback for themselves, to contribute to improve training methods which may benefit others and themselves, and support research performed in the department. In study II, we included participants from two other hospitals in the Oslo region in addition to acquire sufficient participants. These participants were also all positive. We therefore assess that we procured a representative participation to answer the study objective.

All the participants included in the studies completed all training sessions. The combination of defined time for training, pre-set performance goals, a training programme and a trainer might have contributed to commitment to finish all sessions in study II, and the training programme in study III. We do not know if more unstructured training by themselves would have made a difference, but a previous systematic review of voluntary participation in skills training has shown that unrestricted access to simulator equipment is not enough to make trainees practice (69). Furthermore, Strandbygaard and co-authors had dropouts on the different steps in their curriculum study (25). They explained that the main reason was likely to be decreased motivation due to no defined training time.

In previous studies, an intermediate group in terms of surgical experience has been included, and we did the same in study I and II (70) (71). The results of this group, when it came to surgical skills, were not statistically significant different for some tasks, compared to the unexperienced group, and in other tasks, compared to the experienced group. This was also the results in the previous studies. For practical reasons, the most interesting difference, as we see it, is between the inexperienced and the experienced group. The inexperienced are the target group for the training and represent a base from where we register improvement. The experienced performance is the target level for proficiency. The intermediate experienced are the group in the middle, on their way to becoming experienced, representing inexperienced with some experience, they have a learning curve as everyone else, and consequently only add information we already know. With training you gradually perform better and are approaching the proficiency goal.

Nevertheless, the results of the intermediate experienced group were of interest in study I, since it can be used to evaluate if training to a proficiency level before surgery can result in a steeper learning curve like Larsen et al. have demonstrated (60). Konge et al. verified a learning curve on a complex procedure (video assisted thoracoscopic surgery) and how it can be affected with training (72). Inexperienced trainees with training perform as the intermediate experienced group in the clinical setting. This is important knowledge when it comes to acquiring skills when the number of available patients/procedures is decreasing.
Within the experienced group, the practical skills varied when the study participants were assessed using GOALS and CAT-LSH during the surgical procedure (LSH) and also on the different tasks on the simulator. In our opinion, this is likely to be caused by a variation of experience within the group. Self-reported data of the number of previous procedures performed is consequently, in our opinion, insufficient to ensure a homogenous group of study participants in terms of surgical competence. Furthermore, the number of previous procedures needed to become an experienced surgeon is dependent of individual characteristics. Consequently, a practical pre-test with inclusion criteria that assesses the actual skills that are measured, should be taken into consideration when aiming to include participants in a group of experienced surgeons who are supposed to be defined as at the proficiency level. Retrospectively we should have evaluated the participants defining the experienced group with objective assessment tools before including them. The skill level necessary to be included into the experienced group could have been discussed with colleagues selected to work with advanced laparoscopy who have completed curricula like GESEA and were interested in education. We have searched, but not identified, any previous studies that have used a pre-test to evaluate if the participants are good enough to represent the experienced group. Pre-tests are used in studies to evaluate the effect of intervention; like completing a curriculum, training period, a number of repetitions (73).
High volume surgeons are likely to have, but without a guarantee, a high proficiency level of good surgical skills. Other factors are likely to affect the outcome, including paying attention to every detailed step in the procedure and the set-up of the surgical training. By this we mean that they practice, and participate in the training programmes at the department and make themselves available for evaluation. Effort to decrease heterogeneity and variation among surgeons should be accounted for when conducting comparative analyses between surgical techniques (74). Cook et al. point out that confounding in the expert-novice comparison must be taken into consideration (34). There are explanations for observed differences between the groups. The absence of hypothesized differences would suggest a serious flaw in the validity

argument (34). In future studies hypothesized differences between the groups must be discussed when planning the studies and presenting the results.

Assessment tools.

Previous studies have provided evidence of GOALS as an assessment tool for objective measuring of technical skills in basic laparoscopic procedures, which can distinguish surgeons of varying skill levels (75, 76). We replicated this, when comparing unexperienced and experienced participants performing a laparoscopic supracervical hysterectomy procedure. The difference in GOALS scores was not statistically significant for the intermediate experienced and the experienced participants, when assessed by the blinded observers. An explanation can be that GOALS measures fundamental laparoscopic skills, and the intermediate and experienced participants had fundamental skills to an extend that made them equal as a group.

Autonomy is an item of GOALS that might be difficult to assess on a video-recorded procedure. We chose to include all six items of GOALS since we compared the scores between the blinded observers and the surgical assistant. There was no sound on the video recordings, so the blinded assessors did not have access to any verbal guidance or asked questions from the participants. Assistant’s intruments were visible in the videos. The assessors would notice guidance and involvement and take that into consideration during the evaluation of the different steps of the operation when evaluating autonomy. Most of the assessment tools have a grading of autonomy so choosing another tool would not make the video assessment easier. Some studies have kept the autonomy variable, like Vassiliou et al. (77), and others have excluded it from video evaluation like Kramp et al. (24). The latter excluded the item since only parts of the video in which the trainee performed as operating surgeon were edited. Autonomy and good decision making are the target for surgical training, and are therefore essential to evaluate as part of the surgical performance also using video evaluation without sound.

CAT- LSH

There was a statistically significant difference in scores between all groups when using CAT-LSH. To have a procedure specific assessment tool for a procedure (LSH in our studies) provides valuable possibilities, as demonstrated by Mottri et al. (78). They registered that there were specific steps of the prostatectomy, where the scores varied among the experienced surgeons. This part of the procedure was essential for patience outcome, as in potency and continence. A procedure specific assessment tool can be useful by enabling the identification of such steps of a procedure that needs to be improved by a particular surgeon, in order to reduce risk of negative outcomes for the patients.

Assessment of technical skills in surgery should include decision making. The assessment needs to include both the skills and the decision making process (79). With the CAT-LSH, the evaluation of each step of the operation, gave a possibility to ask the participant questions if the surgical strategy was changed. If there was a myoma or an adhesion that affected the strategy and the trainee needed to adapt due to the anatomical conditions, we could have asked why the strategy was changed and how they planned to move on.

Tasks

In the laparoscopic supracervical hysterectomy procedure basic skills are required and in study II, we combined the practical part with the procedural. When evaluating the "Does level" in Miller's pyramid, we should end up evaluating both skills and decision making. This was possible in our study setting in the operation room, but not to the same extent on video recordings, since we had no sound and could not ask the participant what he or she considered as important when changing strategy. This is a limitation of this assessment tool when it comes to video evaluation, but gives advantages in all other situations.

Tasks available on the simulator was essential. In the absence of a basic task, the salpingectomy had the objective of coordinating coagulation and cutting. The participants used a grasper, scissors and bipolar forceps to remove the left tube. The total time used on the task (min) was registered. In the case of an error (bleeding), it had to be corrected before finishing the salpingectomy. It proved appropriate to include the salpingectomy. When we completed the study, it was not usual to include salpingectomy when performing a hysterectomy, as it is today. This means that the curriculum has not lost its topicality.

There was no available laparoscopic supracervical hysterectomy procedure on the simulator when we performed the study, so we chose to use parts of the total laparoscopic hysterectomy procedure for the simulator training in the curriculum. This made the laparoscopic supracervical hysterectomy procedural task incomplete. Essential parts of the operation had to be demonstrated with the actual device, like how to use the loop for the amputation of the cervix, and how to carry out the morcellation. The placement of the loop was marked on the cervix with a hook at the end of the simulation, and the plane for the amputation was evaluated. We considered this sufficient, since decision making in placing the loop for the sake of anatomical structures was demonstrated. In our opinion, it is sufficient to practice part

of a surgical procedure as long as you know that other parts have already been practiced and the participants possess that skill, e.g. entrance and closure.
Suturing is not a part of a standard laparoscopic supracervical hysterectomy, but is an essential skill in case of complications. To incorporate skills necessary for potential complications into a curriculum might give the trainees a feeling of security entering the operation room, since they have skills beyond the actual standard procedure. Retrospectively we should have included suturing since suturing is part of basic skills in laparoscopy, but not a skill every gynaecologist doing laparoscopy possesses.

Rating scale

We developed a rating scale that made it possible to assess and combine the different skills in one total score. This was convenient since the trainees could have one learning curve for the task, not one for each skill practiced in the task. At the same time, they knew which component of the score that was insufficient, and could improve that particular skill.
Another consequence of putting together a total score was to focus on other parameters than time. There is an ideal time for every surgical procedure on the simulator, and we should focus on that, not on speed. In our experience, there is a tendency of not paying sufficient attention to tissue handling and haemostasis if the procedure is performed too fast. We know that infections after surgery often are a result of insufficient haemostasis and hematomas (80, 81). On the other hand, if the procedure takes too long, we have experienced that the gynaecologist/trainee often is not sufficiently familiar with the procedure and/or does not have sufficient practical skills. Unnecessary long procedure time also increases the risk of complications, such as postoperative infections and venous thrombosis (80). When planning the study we considered the rating of the different skills in each task, and whether they were essential to assess and evaluate proficiency, e.g. tissue handling, errors, and hand movements. We registered different parameters in the five tasks, the total score was consequently different in the different tasks. By assessing parameters such as tissue handling like traction and bleeding more than time, we educated the participants to pay attention to the tissue, the surgical strategy and they thus established good habits.
The rating scale for the different parameters registered on the simulator was from 0-2. At the beginning of the learning curve, it was possible to get 0 points. It might be demotivating to get 0 points. We did not ask the trainees about it, and therefore we do not know. But in future

studies, a possible demotivating effect of 0 points can easily be avoided by setting the lowest possible score to be 1.

The rating scale was used on repeated tasks performed by the same participants. It consequently provided longitudinal and continuous information, and thus formed a learning curve, useful both to maintain the participants' motivation and for formative feedback.

Training sessions

The optimal training time and distribution is unknown. Distribution of training has shown to be an effective instructional method (26, (82). This principle has been studied and compared to mass training within the field of laparoscopy. The results are in favour of recommending distributed training sessions in terms of improved training outcomes (28, (83).

Our study design did not further elucidate this issue. However, laparoscopic simulation training has proven to be effective in developing skills, but is a challenge to integrate into work-hour restrictions and non-optimal practice schedules (84). Studies of home training, recommended training schedules from two timed 30 min per week to 1 h per day (85). For training at in hospital the recommendations are almost the same (86).

There was a relatively large variation in the total training period in study II and III. Retrospectively, we might have planned for training sessions with a fixed frequency, to have equal distribution of training for all the study participants. This could have been feasible in a study setting, but the validity of the results would, in our opinion, be reduced as this is difficult to plan and carry out in a clinical setting. Maintenance training is a valuable and necessary addendum to proficiency-based training programmes for laparoscopic suturing. A maintenance training interval of 1 month with unsupervised training sessions on box trainers seems ideal (84).

Steps in the curriculum

A theoretical knowledge part, followed by a practical part, before the participants move on to the operation room, are recommended steps of a curriculum, and have also been studied by others (20, 25, 87). The theoretical part in our study was a written test. We could have extended the theoretical part with reading material and procedural video tutorials to improve the knowledge and their understanding of the tasks and procedures in the practical part. Available textbooks, journal articles and available videos from for WebSurg could have been used. Retrospectively, we might have highlighted the steps of the LSH in the theoretical part, and made sure that the trainees understood the steps of the surgical procedure, why these steps were a good strategy, and prepared them to handle unexpected events. We introduced the strategy in step 2, as part of the instruction for the hysterectomy task on the simulator. Kohls-Gatzoulis et al. point out that success of a surgical procedure depends on more than the ability to perform each of the manoeuvres associated with the procedure (88). Cognitive skills such as error detection, forward planning and decision making are essential. Surgeons need to be trained to judge the correctness of their operation (88). Tang et al. studied errors made by surgical trainees during skills training courses, and demonstrated that cognitive errors, like lack of understanding the correct sequence of steps in an operation trigger the majority of mistakes in a procedural task rather than technical errors (40). This finding supports the importance of self-assessment. We could have let the participants evaluate themselves after surgery or the video recording of the performed procedure, and presented it to a supervisor afterwards.

The curriculum could furthermore have been improved with implementation of morcellation in the theoretical part (step 1) and in the practical part (step 2). This would have made the training sessions more time consuming but presented the LSH procedure as a whole. Kurashima et al. built a training model for the entire procedure of a laparoscopic inguinal hernia repair and used a procedure specific rating scale GOALS-GH (89). They included evaluation of trocar location and placement, justified by a competent surgeon who was expected to have accurate understanding of trocar selection and how to modify placement for different patients related to anatomy and practice of safe trocar insertion techniques with good visualization of accessory trocar entry. This could also have been included in the evaluated procedure in our study.

The participants in our study performed the theoretical test at home. Implementation of a traditional exam performed at work would possibly have motivated the candidates for better theoretical preparation before the test. There are strong indications that students prioritise the knowledge and skills that they know may be included in the exam, particularly when a grading scale is used (90). It has also been shown that testing can enhance motivation, concentration, and effort (91). Therefore, it is not recommended to remove the theoretical part of a similar curriculum in the future, rather it could be expanded it in future studies.
We went through the results of the theoretical test with each candidate, and wrong answers were discussed to enable the trainee to give the correct answer. We could also have reviewed the correct answers, and highlighted why it is important to make sure they had the necessary

knowledge and not obtained a right answer by guessing. However, this would have been more time consuming.

We chose a validated questionnaire in basic laparoscopy containing 34 multiple-choice questions. Multiple choice questions however, have limitations and strengths. Students are not necessarily tested on their ability to form a sensible opinion, but on choosing between response alternatives. Ideally the questions should test the candidates' understanding and ability to apply knowledge, but we would end up with questions that instead test superficial knowledge and ability to reproduce knowledge (90). Such questions are at a lower step of the learning pyramid.

Moreover, previous studies indicate that multiple choice questions have an unfortunate effect on the students' learning behaviour and mindset (90). By reading previous exam questions in advance, the students learn to recognise correct and wrong answers to questions. They learn how to guess by looking for clues in the response alternatives. Such questions may thus give a false impression of the students' skills and are a poor test of the ability to think, to act independently and to respond to an academic challenge. The chance of answering correctly by guessing increases if one or two of the of four questions are formulated in such a way that even weak candidates realise that they are incorrect (90). Some students answer correctly only after recognising the correct answer from among the response alternatives, described in the academic literature as the "cueing effect"(90). Multiple choice questions also have advantages. In a test, the questions can be answered quickly and therefore the test may contain a high number of questions and thereby cover more subjects in the relevant discipline. Studies show that multiple choice questions can distinguish relatively well between strong, average and weak candidates (90). Multiple choice questions are furthermore easy to correct, and thereby faster to administer compared to free-text questions. Agreement about what is the right answer makes it fair for the students. We included 6 free-text questions in the test to adjust the theoretical part of the LSH. A disadvantage with free-text questions can be that they are difficult to score in an objective manner. Consistent scoring practice can be secured by a scoring template, pilot answers before the exam, and by having two sensors that confer with each other. It takes longer to answer and thus results in that the number of questions are lower than if an exam contains only multiple choice questions.

Procedure

Multiple surgical procedures, such as appendectomy, cholecystectomy, nephrectomy, and hysterectomy, are available on surgical simulator systems. The laparoscopic supracervical hysterectomy was, when the study was conducted, the procedure preferred in our department for trainees who were about to start performing more advanced laparoscopic procedures. We could have made the curriculum with another elective procedure considered easier to perform like salpingo-oophorectomy or more complex like the total laparoscopic hysterectomy. This would have influenced the tasks included in the curriculum, and the time to complete the tasks. For total laparoscopic hysterectomy, the time most likely would have been increased, since suturing had to be a part of the practical training. The crucial reason for the choice of procedure was that the LSH was, and still is, a good choice for selected women who need a hysterectomy due to a benign condition. Furthermore, a large number of laparoscopic supracervical hysterectomy procedures were performed at our department, which facilitated the implementation of our studies.

Evaluation of surgical performance using assessment tools.

Blinded assessors might reduce possible bias when evaluating colleagues, friends, and knowing their level of competence (92). It was of importance to reduce possible bias since we compared GOALS as an already validated assessment tool, to CAT-LSH, that had never been evaluated before. Furthermore, video evaluation has some practical advantages as the evaluation can be performed at any suitable time/hour. This flexibility made it easier for us to have the same laparoscopic experts evaluating all videotaped procedures in study I and III. We chose to include assessment performed by the assistant surgeons in addition, because it was of interest to register how GOALS and CAT-LSH worked out in the situation similar to the clinical setting where the assessment tools were supposed to be used. This evaluation was done regardless of whether the assistant was a trainee or an experienced gynaecologist. It is previously shown that both residents and experienced physicians can assess surgical skills (93). We could have omitted assessment either by the blinded observers or the assistant surgeons and still found the same results when comparing inexperienced and experienced study participants. This would have saved us time, but on the other hand, since CAT-LSH had not been previously validated, we wanted concurrent validity from different assessment situations. Furthermore, the assessment from the assisting surgeons might be the easiest feedback to implement in a clinical setting, and the results were therefore of interest.

To let the trainees evaluate themselves was also a possibility. Studies of self- assessment gives results that consistently do not differ from expert assessments (94). We considered this possibility, but concluded that we had sufficient data with the registration from the two blinded assessors and the assisting surgeon. Retrospectively we might have included self-assessment, since surgeons need to be trained to judge the correctness of their operations. For the future we would recommend self- assessment as part of the training, for example by letting the trainees evaluate themselves and afterwards discuss the assessment with a supervisor. The blinded observers in our studies were both experts in laparoscopy and familiar with assessing trainees. They were instructed of how perform assessment using GOALS and CAT-LSH, prior to the assessment of the study participants performance. As commented in previous studies, inexperienced assessors have a tendency of not using the whole rating scale. This is known as central tendency bias, or end aversion bias in the relevant literature.(95). Experience with the assessment tool may be necessary to enable reliable skill assessment (96). Consequently, a reduction of the number of assisting surgeons to one or two might have influenced the assessment score. We did not identify whether there was a different understanding of what the assessors defined as good and not so good laparoscopic skills among the assisting surgeons in our study.

We found that the overall interclass correlation coefficient was good for all comparisons (> 0.7). Also the interrater reliability scores of the two blinded observers were good for both GOALS and CAT-LSH. Furthermore, when comparing the blinded observers and the surgical assistants the interrater scores were satisfactory.

9.2 Interpretation of results

Our results add to the results of many previous studies, that show that training can be structured, feedback can be objective and trainees can be prepared for operations in the operation theatre

Our studies did not result in chance in practice, but it can contribute to a change. The design and results of our studies might give opportunities and motivation to trainees, supervisors and the employers to implement formative and summative assessment in laparoscopic surgery.

9.3 General consideration

There are limitations and strengths with proficiency-based education, assessment tools and simulator training. Our studies, as well as many others, provide knowledge on how to assess, evaluate and implement proficiency-based education. Some of our results are specific for training in gynaecology, but other aspects are transferable for training and assessment of surgical skills in general.

Strengths

The curriculum had a theoretical and a practical part with training goals. The training and the evaluation using the assessment tools, took place in a clinical setting, not in a simulated one. This makes the test situation realistic and the results are consequently most likely to be transferable to clinical practise.

Different learning approaches and frameworks have been described and are useful when it comes to planning, implementing and evaluating clinical competence. According to Miller's pyramid of competence development, GOALS and CAT- LSH give the possibility to evaluate at the highest level of competence; the does level (45).

The sample size was a result of the power calculation and must therefore be considered sufficient to answer the study questions. It can be a waste of time for the participants, and extra work for the persons running the studies, if more participants than necessary are included.

It is a strength of our studies that all participants included in the studies completed all the training sessions. In case we had had drop-outs, there would have been the risk of a selection bias. The combination of defined time for training, pre-set performance goals, a training programme and a trainer might have contributed to commitment to finish all sessions in study II, and the training programme in study III.

The CAT-LSH gives a possibility for formative feedback. The scoring system in study II can potentially be used on all types of procedures/tasks as long as measurable parameters are available. The scoring system combines parameters like time, movements, errors, etc. and gives a total score. We tested it at repeated sessions and showed that it can give a learning curve, and thereby, formative feedback.

Limitations

Study I and III were single-centre studies. We calculated the sample size in order to decide the necessary number of study participants to answer the study questions. Recruitment of study participants from different departments in all three studies might have had an impact on the result, as our department had a high rate of laparoscopic surgery.
The outcome for the patients in study III was not registered. Johnstone et al. suggested that desired health outcomes including the avoidance of mortality, complications and prolonged length of stay can be measured and used to answer whether simulation-based training can improve patient safety, when defined as reduction of harm to patients (97). Given the small sample size, it is unlikely that such measures would have added anything in our studies. The patients included in the studies did not have any special follow up after the surgery. They were as all patients at our department, and registered in the Norwegian Gynecological Endoscopic Register, which is a national register for outcomes after gynaecological endoscopic surgery.

We reported the results of the different tasks in study II as a total score. We could have added the benchmark metrics for the different tasks so others that have the same simulator with the same tasks as we used, could use the results if they wanted to. It is only the principle of the scoring system that is transferable to others, not a scoring system ready for use. On the other hand, one study is not enough to decide a score for summative feedback applicable for a national standard. Regardless of that it would have been a possible start for those having the same simulator as the one we used in our studies.

The simulated LSH procedure was not complete since it did not include the application of the loop and the morcellator. There was no complete procedure available, and with adjustments we considered it good enough. Before the actual procedure the loop was shown to the participants and they tested the principle with loosening and tightening. The simulated LSH included a marking on the cervix where to place the loop, so anatomic knowledge was tested. The morcellator was demonstrated for the candidate before surgery, in step 3 in study III. The lack of training in these two particular aspects of the procedure before commencing in the operation room is an obvious limitation of our study. To compensate for this, we could have considered a special focus on these particular parts of the procedure using videos during the theoretical part of the curriculum. However, many specific procedures are not included in the different simulators. We consequently will have to improvise different procedures and accept that the face validity might be rated as low. Stefanides et al. have experienced that video tutorials shown during simulation training might influence the achievement of skills in a positive way (98). We used no video tutorial of the LSH procedure, which could have improved the performance of our study.

We did not discuss if the results of the experienced participants could have been used as a cut off score to set a proficiency level used in summative feedback for the LSH in the OR. When do trainees have sufficient competence do a procedure by themselves, or with a less experienced assistant? This has been carried out by Jelkovsek et al. (99). They established performance cut off scores for vaginal hysterectomy (51).

For the future we can hope for a national standard that gives guidelines with suggestions and advice on available curricula and assessment possibilities with performance cut-off scores. It is our hope that our studies and this thesis will contribute to a development towards a more systematically approach for training within surgery.
Our studies contributed to individual skills for the study participants, but for optimal surgical performance individual technical skills are not enough. Team-based skills like communication, situation awareness and decision making are necessary and must be practised (100). To achieve this, the team that work together in the operation room, must practice together.

9.4 Added value of the performed studies

1. The “sign them in” principle turns the focus towards the system not on the individual gynaecologist. The employers have the final responsibility that the gynaecologists in the team have the necessary surgical skills.

2. The trainees and gynaecologists know that they are good enough for their patients.

3. The “sign them in” principle gives the trainees a clear goal for their training and the assessment is objective and thereby fair.

4. Development and validation of a procedure specific assessment tool for laparoscopic supracervical hysterectomy, CAT-LSH

5. Development and validation of a rating scale for proficiency. Potentially it can be used on all types of procedures/tasks as long as measurable parameters are available. The scoring system combines parameters like time, movements, errors, etc. and gives a total score. We tested it at repeated sessions and showed that it can give a learning curve, and thereby, formative feedback.

6. An example of a curriculum for the LSH, with registered time needed to reach the proficiency-levels.

10. Conclusion and future perspectives

10.1 Conclusions

Through the three papers forming the basis of this PhD thesis, we have shown that we can measure surgical skills by using GOALS and CAT-LSH as assessment tools for laparoscopic supracervical hysterectomy. We developed and validated a scoring system that can be used for summative and formative feedback in proficiency-based assessment. With this or similar systems, it is possible and useful to customize parameters with difference designation into a total score. In study III, we have shown, as many previous authors also have, that accomplishment of a curriculum with proficiency levels on selected tasks on a simulator, enables trainees to perform surgery at a more advanced level, defined as higher scores on a LSH performance, assessed with CAT-LSH and GOALS, than trainees without preoperative skill training.

10.2 Future perspectives

Changes in medical treatment have decreased surgical opportunities for residents. Simulation based training should be regarded as an adjunct to the apprenticeship-based model in the OR and can partly solve challenges related to reduced surgical exposure.

This raises relevant questions such as:

-What knowledge, skills and abilities are needed and what required education is necessary for gynaecologists today?

-What are the core competencies that all trainees must learn whether they end up in specialist practice outside the hospital, or as a generalist in a small hospital department or a subspecialist in a university hospital?

Defining the minimum standard for all ob-gyn residents is not the same as defining the range of competencies required of a specialist. There is a core-set of knowledge, skills and abilities that all gynaecologists need to know. Subspecialists have additional skills and knowledge. The variety of technical complexity of procedures has expanded (101). In gynaecology, we currently have four main modalities for hysterectomy; laparotomy, the vaginal route, laparoscopy and robot assisted. This development results in a challenge in the education concerning available procedures for the trainees. Can simulated procedures like LSH and total laparoscopic hysterectomy replace real operations for the trainees, and leave a higher volume to the gynaecologists to maintain competence and educate successors? Larsen et al. showed that trainees performing a salpingectomy for the first time after systematic training, operated with the same evaluation score as colleagues who had previously performed 30-50 procedures (102). Driessen et al. demonstrated in a publication entitled “Proficiency for Advanced Laparoscopic Procedure in Gynaecologic Residency Program: Do all Residents Need to be Trained” that 56% of the asked gynaecologists did not perform any level 3 procedures (hysterectomy, myomectomy, extensive adhesiolysis, and severe endometriosis), and 86 did not perform level 4 procedures (sacrocolpopexy, lymphadenecomy, and recto vaginal endometriosis) (103). We must know the difference between training in the operation room and training in the skills lab. A trainee can learn to suture on a simulator, but that does not make her ready to perform a procedure in the operation room (104). Situated learning, where relational and social aspects about how to become and act as part of a team is missing if we only have technical skills training outside the operation room (104). With implementation of duty-hour regulations, concerns were raised that trainees might not have adequate time to develop competencies in the required surgical skills. The reduction of in-house hospital experience also includes a decrease in opportunities for assisting in surgeries, which is an important component of surgical training (105). Simulation can shorten the learning curve when it comes to practical skills, but cannot replace the learning of behaviour in the operation room. We think the solution is a combination. The trainees can simulate several hysterectomies to a pre-set proficiency level, and then assist/operate a number of hysterectomies and be evaluated with an objective evaluation tool.

Proficiency based surgical education is an advantage for patients, doctors and employers. It protects the patients from doctors who have not yet reached proficiency (106). It is important to remember that training in the operation room does not necessarily place the patient at a higher risk, because the trainee is under close supervision by the supervisor. With proficiency-based education, the trainees gain summative and formative objective feedback, and know what is expected of them when it comes to surgical skills. It is transparent and fair since you need to be signed in before you carry out surgery, and the result is fairness among the trainees. We know from experience that practical training and evaluation is viewed as useful, and feedback is something the trainees want, and miss when it is lacking in their education (90).

The assessment tools and scoring system on the simulator is predictable, objective and provides feedback immediately after the performance. Robotic surgery is increasing within the field of gynaecological surgery, and might take over as the main surgical technique in the future. The results of previous publications indicate that also robotic skills can be aided by laparoscopic training (107-109).

For employers systematic training and certification is a quality assurance since in an objective way they can register that the employees have the necessary qualifications. There is no evidence whether the number requested in a logbook reproduces the skills acquired by the resident (3).

Proficiency based training and evaluation can be a part of the culture and tradition in the department. The employers can make it compulsory in the education plan, and ensure implementation. It affects the learning environment in a positive way. When it is compulsory, it is not stigmatizing. Practical training and evaluation are aspects that everyone comprehends, and are not something some employers need when they have made a mistake, a patient has complained or there is a supervisory case. Furthermore, in supervisory cases, it is of importance that employers can document that "your doctors" have been signed in to surgery and have the necessary skills. Trainees do not voluntarily participate in simulation-based laparoscopic skills training (110). It is necessary to define proficiency levels and expect trainees to have accordant basic skills before they operate on patients. The investments the hospital have made in equipment in a skills lab is probably only effective when practicing outside the operation room is a mandatory part of the curriculum. With the demands on efficiency in the operation room, the apprenticeship model is under pressure. Given that simulation-based education can replace training in the operation theatre, it is a good idea to move training into the skills lab. Trainees can prepare on simulators before surgery, and thereby reduce operating time, which again results in opportunities for better utilization of the operation theatre.

The hospitals provide us with specialised treatment. In addition, the hospitals have tasks involving research and training, as well as educating patients and their relatives.
The Ministry of Health and Care Services has the supervisory responsibility. In order to contribute to education and training in surgery, we have a responsibility to inform The

Ministry of Health and Care Service about what we do, how we do it and the cost. Safer surgeons faster, summarizes the benefits with training. Sign them in, summarizes the way to achieve it.

11. Reference list

1. Wulf G, Shea C, Lewthwaite R. Motor skill learning and performance: a review of influential factors. Med Educ. 2010;44(1):75-84.
2. Katarjian. Relevance of the Hippocratic Oath in the 21st Century. The ASCO post. 2014.
3. Fritz T, Stachel N, Braun BJ. Evidence in surgical training - a review. Innovative surgical sciences. 2019;4(1):7-13.
4. Cameron JL. William Stewart Halsted. Our surgical heritage. Ann Surg. 1997;225(5):445-58.
5. Harvey AM. The influence of William Stewart Halsted's concepts of surgical training. Johns Hopkins Med J. 1981;148(5):215-36.
6. Kerr B, O'Leary JP. The training of the surgeon: Dr. Halsted's greatest legacy. Am Surg. 1999;65(11):1101-2.
7. Hamdorf JM, Hall JC. Acquiring surgical skills. Br J Surg. 2000;87(1):28-37.
8. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004;79(10 Suppl):S70-81.
9. Litynski GS. Kurt Semm and the fight against skepticism: endoscopic hemostasis, laparoscopic appendectomy, and Semm's impact on the "laparoscopic revolution". JSLS. 1998;2(3):309-13.
10. Bhattacharya K. Kurt Semm: A laparoscopic crusader. J Minim Access Surg. 2007;3(1):35-6.
11. Mettler L, Eckmann-Scholz C, Semm I, Alkatout I. Factors to consider in gynecological surgery. Women's health (London, England). 2014;10(3):323-38.
12. Langebrekke A, Skår OJ, Urnes A. Laparoscopic hysterectomy. Initial experience. Acta Obstet Gynecol Scand. 1992;71(3):226-9.
13. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc. 2003;17(10):1525-9.
14. Gynecology TNFoSoGoOa. Clinical Guidelines in Gynecological Endoscopy. 1999.
15. Dictionary HS. [Available from: https://www.ahrq.gov/sites/default/files/wysiwyg/patient-safety/resources/simulation/sim-dictionary-2nd.pdf.
16. Estai M, Bunt S. Best teaching practices in anatomy education: A critical review. Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft. 2016;208:151-7.
17. Martelli N, Serrano C, van den Brink H, Pineau J, Prognon P, Borget I, et al. Advantages and disadvantages of 3-dimensional printing in surgery: A systematic review. Surgery. 2016;159(6):1485-500.
18. Willaert WI, Aggarwal R, Van Herzeele I, Cheshire NJ, Vermassen FE. Recent advancements in medical simulation: patient-specific virtual reality simulation. World J Surg. 2012;36(7):1703-12.
19. Hiemstra E, Kolkman W, Jansen FW. Skills training in minimally invasive surgery in Dutch obstetrics and gynecology residency curriculum. Gynecol Surg. 2008;5(4):321-5.

20. Stefanidis D, Heniford BT. The formula for a successful laparoscopic skills curriculum. Arch Surg. 2009;144(1):77-82; discussion
21. Fried GM. Lessons from the surgical experience with simulators: incorporation into training and utilization in determining competency. Gastrointest Endosc Clin N Am. 2006;16(3):425-34.
22. Campo R, Wattiez A, Tanos V, Di Spiezio Sardo A, Grimbizis G, Wallwiener D, et al. Gynaecological endoscopic surgical education and assessment. A diploma programme in gynaecological endoscopic surgery. Gynecol Surg. 2016;13:133-7.
23. Gesea Educational Program. GESEA MIGS 2022 [Available from: https://gesea.eu/pathways/.
24. Ritter EM, Scott DJ. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. Surg Innov. 2007;14(2):107-12.
25. Strandbygaard J, Bjerrum F, Maagaard M, Rifbjerg Larsen C, Ottesen B, Sorensen JL. A structured four-step curriculum in basic laparoscopy: development and validation. Acta Obstet Gynecol Scand. 2014;93(4):359-66.
26. Shore EM, Grantcharov TP, Husslein H, Shirreff L, Dedy NJ, McDermott CD, et al. Validating a standardized laparoscopy curriculum for gynecology residents: a randomized controlled trial. Am J Obstet Gynecol. 2016;215(2):204.e1-.e11.
27. Moulton CA, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. Ann Surg. 2006;244(3):400-9.
28. Lee TD, Genovese ED. Distribution of practice in motor skill acquisition: different effects for discrete and continuous tasks. Res Q Exerc Sport. 1989;60(1):59-65.
29. Mitchell EL, Lee DY, Sevdalis N, Partsafas AW, Landry GJ, Liem TK, et al. Evaluation of distributed practice schedules on retention of a newly acquired surgical skill: a randomized trial. Am J Surg. 2011;201(1):31-9.
30. Brown HL, Carson SA, Lawrence HC, 3rd. The First National Summit on Women's Health: The Future of Obstetrics and Gynecology Training. Obstet Gynecol. 2018;132(3):755-62.
31. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357(9260):945-9.
32. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ. 2003;37(9):830-7.
33. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 2015;49(6):560-75.
34. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119(2):166.e7-16.
35. Berg M. Cutting Edges in Surgical Training. In: Mollo RAaEA, editor. 1: Via Media; 2013. p. 42.
36. Witheridge A, Ferns G, Scott-Smith W. Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. International journal of medical education. 2019;10:191-2.
37. McKendy KM, Watanabe Y, Lee L, Bilgic E, Enani G, Feldman LS, et al. Perioperative feedback in surgical training: A systematic review. Am J Surg. 2017;214(1):117-26.
38. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273-8.

39. Gjersvik P. Why is feedback so difficult? Tidsskr Nor Laegeforen. 2014;134(6):591.
40. Sidhu RS, Grober ED, Musselman LJ, Reznick RK. Assessing competency in surgery: where to begin? Surgery. 2004;135(1):6-20.
41. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ. 2003;327(7422):1032-7.
42. Reznick RK. Teaching and testing technical skills. Am J Surg. 1993;165(3):358-61.
43. Grantcharov TP, Schulze S, Kristiansen VB. The impact of objective assessment and constructive feedback on improvement of laparoscopic performance in the operating room. Surg Endosc. 2007;21(12):2240-3.
44. Soper NJ, Fried GM. The fundamentals of laparoscopic surgery: its time has come. Bull Am Coll Surg. 2008;93(9):30-2.
45. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9 Suppl):S63-7.
46. Teasdale GM. Learning from Bristol: report of the public inquiry into children's heart surgery at Bristol Royal Infirmary 1984-1995. Br J Neurosurg. 2002;16(3):211-6.
47. Van Der Weyden MB. The Bundaberg Hospital scandal: the need for reform in Queensland and beyond. Med J Aust. 2005;183(6):284-5.
48. Irvine D. The performance of doctors. I: Professionalism and self regulation in a changing world. BMJ. 1997;314(7093):1540-2.
49. Lim TO, Soraya A, Ding LM, Morad Z. Assessing doctors' competence: application of CUSUM technique in monitoring doctors' performance. Int J Qual Health Care. 2002;14(3):251-8.
50. Buanes T. Advarer mot utrente kirurger. VG.
51. Campo R, Molinas CR, De Wilde RL, Brolmann H, Brucker S, Mencaglia L, et al. Are you good enough for your patients? The European certification model in laparoscopic surgery. Facts, views & vision in ObGyn. 2012;4(2):95-101.
52. Gibbs VC. Thinking in three's: changing surgical patient safety practices in the complex modern operating room. World journal of gastroenterology. 2012;18(46):6712-9.
53. Soot SJ, Eshraghi N, Farahmand M, Sheppard BC, Deveney CW. Transition from open to laparoscopic fundoplication: the learning curve. Archives of surgery (Chicago, Ill : 1960). 1999;134(3):278-81; discussion 82.
54. Tanos V, Socolov R, Demetriou P, Kyprianou M, Watrelot A, Van Belle Y, et al. Implementation of minimal invasive gynaecological surgery certification will challenge gynaecologists with new legal and ethical issues. Facts, views & vision in ObGyn. 2016;8(2):111-8.
55. Davis NL, Willis CE. A new metric for continuing medical education credit. J Contin Educ Health Prof. 2004;24(3):139-44.
56. Kumta SM. Training Junior Surgical Residents to Attain Mastery. In: Berg M, editor. Cutting edges in surgical training2013. p. 27-31.
57. Schijven MP, Berlage JT, Jakimowicz JJ. Minimal-access surgery training in the Netherlands: a survey among residents-in-training for general surgery. Surg Endosc. 2004;18(12):1805-14.
58. Kolkman W, Wolterbeek R, Jansen FW. Gynecological laparoscopy in residency training program: Dutch perspectives. Surg Endosc. 2005;19(11):1498-502.
59. Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. Surgery. 2005;138(4):640-7; discussion 7-9.

60. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ. 2009;338:b1802.
61. Stefanidis D, Acker CE, Swiderski D, Heniford BT, Greene FL. Challenges during the implementation of a laparoscopic skills curriculum in a busy general surgery residency program. J Surg Educ. 2008;65(1):4-7.
62. Carlsen ASaKCL. The Research Handbook 2021, 9th Edition, revised October 2021. 2021.
63. Miskovic D, Ni M, Wyles SM, Kennedy RH, Francis NK, Parvaiz A, et al. Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England. Ann Surg. 2013;257(3):476-82.
64. Goderstad JM, Sandvik L, Fosse E, Lieng M. Assessment of Surgical Competence: Development and Validation of Rating Scales Used for Laparoscopic Supracervical Hysterectomy. Journal of surgical education. 2016;73(4):600-8.
65. Larsen CR, Grantcharov T, Aggarwal R, Tully A, Sorensen JL, Dalsgaard T, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surgical endoscopy. 2006;20(9):1460-6.
66. Strandbygaard J, Maagaard M, Larsen CR, Schouenborg L, Ottosen C, Ringsted C, et al. Development and validation of a theoretical test in basic laparoscopy. Surg Endosc. 2013;27(4):1353-9.
67. Fagerland MW, Sandvik L. Performance of five two-sample location tests for skewed distributions with unequal variances. Contemporary clinical trials. 2009;30(5):490-6.
68. World Medical Association. WMA Declaration of Helsinki- Ethical principles for medical research involving human subjects 2022 [Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.
69. Gostlow H, Marlow N, Babidge W, Maddern G. Systematic Review of Voluntary Participation in Simulation-Based Laparoscopic Skills Training: Motivators and Barriers for Surgical Trainee Attendance. J Surg Educ. 2017;74(2):306-18.
70. Larsen CR, Grantcharov T, Aggarwal R, Tully A, Sørensen JL, Dalsgaard T, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc. 2006;20(9):1460-6.
71. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. BJOG. 2008;115(7):908-16.
72. Jensen K, Petersen RH, Hansen HJ, Walker W, Pedersen JH, Konge L. A novel assessment tool for evaluating competence in video-assisted thoracoscopic surgery lobectomy. Surg Endosc. 2018;32(10):4173-82.
73. Mansoor SM, Tunold JA, Næss PA, Trondsen E, Gaarder C, Skattum J. Course in basic surgical skills. Tidsskr Nor Laegeforen. 2014;134(9):935-7.
74. Nyberg M, Sjoberg DD, Carlsson SV, Wilderäng U, Carlsson S, Stranne J, et al. Surgeon heterogeneity significantly affects functional and oncological outcomes after radical prostatectomy in the Swedish LAPPRO trial. BJU Int. 2021;127(3):361-8.
75. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2005;190(1):107-13.

76. Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. J Am Coll Surg. 2007;204(2):308-13.
77. Vassiliou MC, Feldman LS, Fraser SA, Charlebois P, Chaudhury P, Stanbridge DD, et al. Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. Surg Innov. 2007;14(3):211-6.
78. Mottrie A, Mazzone E, Wiklund P, Graefen M, Collins JW, De Groote R, et al. Objective assessment of intraoperative skills for robot-assisted radical prostatectomy (RARP): results from the ERUS Scientific and Educational Working Groups Metrics Initiative. BJU Int. 2020.
79. Sharma S. Objective assessment of technical skills in surgery: assessment should include decision making. BMJ. 2004;328(7436):403.
80. Clarke-Pearson DL, Geller EJ. Complications of hysterectomy. Obstet Gynecol. 2013;121(3):654-73.
81. Ravlo M, Moen MH, Bukholm IRK, Lieng M, Vanky E. Ureteric injuries during hysterectomy-A Norwegian retrospective study of occurrence and claims for compensation over an 11-year period. Acta Obstet Gynecol Scand. 2022;101(1):68-76.
82. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, Kunz R. Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine. BMJ. 2002;325(7376):1338-41.
83. Koen W.van Dongen PJ, Mitra, Marlies P.Schijven, Ivo A.M.J. Broeders. Distributed versus Masses Training: Efficiency of Training Psychomotor Skills. Surg Tech 2011.
84. Van Bruwaene S, Schijven MP, Miserez M. Maintenance training for laparoscopic suturing: the quest for the perfect timing and training model: a randomized trial. Surg Endosc. 2013;27(10):3823-9.
85. Gue S. Home-made videoscopic trainer for operative laparoscopic surgery. Aust N Z J Surg. 1995;65(11):820-1.
86. De Win G, Van Bruwaene S, De Ridder D, Miserez M. The optimal frequency of endoscopic skill labs for training and skill retention on suturing: a randomized controlled trial. J Surg Educ. 2013;70(3):384-93.
87. Shore EM, Lefebvre GG, Husslein H, Bjerrum F, Sorensen JL, Grantcharov TP. Designing a Standardized Laparoscopy Curriculum for Gynecology Residents: A Delphi Approach. J Grad Med Educ. 2015;7(2):197-202.
88. Kohls-Gatzoulis JA, Regehr G, Hutchison C. Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. Can J Surg. 2004;47(4):277-83.
89. Kurashima Y, Feldman L, Al-Sabah S, Kaneva P, Fried G, Vassiliou M. A novel low-cost simulator for laparoscopic inguinal hernia repair. Surg Innov. 2011;18(2):171-5.
90. Gjersvik P. Students' exams and doctors' reality. Tidsskr Nor Laegeforen. 2018;138(10).
91. Larsen DP, Butler AC, Roediger HL, 3rd. Test-enhanced learning in medical education. Med Educ. 2008;42(10):959-66.
92. Vogt VY, Givens VM, Keathley CA, Lipscomb GH, Summitt RL, Jr. Is a resident's score on a videotaped objective structured assessment of technical skills affected by revealing the resident's identity? Am J Obstet Gynecol. 2003;189(3):688-91.
93. Oestergaard J, Larsen CR, Maagaard M, Grantcharov T, Ottesen B, Sorensen JL. Can both residents and chief physicians assess surgical skills? Surg Endosc. 2012;26(7):2054-60.

94. Arora S, Miskovic D, Hull L, Moorthy K, Aggarwal R, Johannsson H, et al. Self vs expert assessment of technical and non-technical skills in high fidelity simulation. Am J Surg. 2011;202(4):500-6.
95. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. J Psychosom Res. 2010;68(4):319-23.
96. Matsuda T, Kanayama H, Ono Y, Kawauchi A, Mizoguchi H, Nakagawa K, et al. Reliability of laparoscopic skills assessment on video: 8-year results of the endoscopic surgical skill qualification system in Japan. J Endourol. 2014;28(11):1374-8.
97. Johnston MJ, Paige JT, Aggarwal R, Stefanidis D, Tsuda S, Khajuria A, et al. An overview of research priorities in surgical simulation: what the literature shows has been achieved during the 21st century and what remains. Am J Surg. 2016;211(1):214-25.
98. Stefanidis D, Korndorffer JR, Jr., Heniford BT, Scott DJ. Limited feedback and video tutorials optimize learning and resource utilization during laparoscopic simulator training. Surgery. 2007;142(2):202-6.
99. Jelovsek JE, Walters MD, Korn A, Klingele C, Zite N, Ridgeway B, et al. Establishing cutoff scores on assessments of surgical skills to determine surgical competence. Am J Obstet Gynecol. 2010;203(1):81.e1-6.
100. Vincent C, Moorthy K, Sarker SK, Chang A, Darzi AW. Systems approaches to surgical quality and safety: from concept to measurement. Ann Surg. 2004;239(4):475-82.
101. Johanson ML, Lieng M. Changes in route of hysterectomy in Norway since introduction of robotic approach. Facts, views & vision in ObGyn. 2021;13(1):35-40.
102. Kolkman W, Wolterbeek R, Jansen FW. Implementation of advanced laparoscopy into daily gynecologic practice: difficulties and solutions. J Minim Invasive Gynecol. 2006;13(1):4-9.
103. Driessen SR, Janse JA, Schreuder HW, Jansen FW. Proficiency for Advanced Laparoscopic Procedures in Gynecologic Residency Program: Do all Residents Need to be Trained? J Surg Educ. 2015;72(5):942-8.
104. Våpenstad C, Buzink SN. Procedural virtual reality simulation in minimally invasive surgery. Surg Endosc. 2013;27(2):364-77.
105. Lewis FR, Klingensmith ME. Issues in general surgery residency training--2012. Ann Surg. 2012;256(4):553-9.
106. Lateef F. Simulation-based learning: Just like the real thing. Journal of emergencies, trauma, and shock. 2010;3(4):348-52.
107. Davila DG, Helm MC, Frelich MJ, Gould JC, Goldblatt MI. Robotic skills can be aided by laparoscopic training. Surg Endosc. 2018;32(6):2683-8.
108. Finnerty BM, Afaneh C, Aronova A, Fahey TJ, 3rd, Zarnegar R. General surgery training and robotics: Are residents improving their skills? Surg Endosc. 2016;30(2):567-73.
109. Davidson A, Shore E, Shirreff L, Shah A, Shah R. Validating an Obstetrics and Gynaecology Longitudinal Integrated Clerkship Curriculum at the University of Toronto: A Four-Year Review. J Obstet Gynaecol Can. 2021;43(3):372-5.
110. van Dongen KW, van der Wal WA, Rinkes IH, Schijven MP, Broeders IA. Virtual reality training for endoscopic surgery: voluntary or obligatory? Surg Endosc. 2008;22(3):664-7.

12. Appendices

Global Operative Assessment in Laparoscopic skills- GOALS

Domains	1	2	3	4	5
Depth perception	Constantly overshooting target, hits backstop, wide swings, slow to correct.		Some overshooting or missing plane but corrects quickly.		Accurately directs instruments in correct plane to target.
Bimanual dexterity	Use of one hand, ignoring non-dominant hand, poor coordination between hands.		Use of both hands but does not optimize interactions between hands to facilitate conduct of operation.		Expertly uses both hands in a complementary manner to provide optimal working exposure.
Efficiency	Uncertain, much wasted effort, many tentative motions, constantly changing focus of operation, or persisting at a task without progress.		Slow, but planned and reasonably organized.		Confident, efficient and safe conduct of operation, maintaining focus on component of procedure until better done approach.
Tissue handling	Rough, tears tissue by excessive traction, injures adjacent structures, poor control of coagulation device, grasper frequently slips off.		Handles tissues reasonably well with some minor trauma to adjacent tissues, e.g. coagulation of non-target tissue, occasional slipping of grasper.		Handles tissue very well with appropriate traction on tissues and negligible injury of adjacent structures. Uses energy sources appropriately but not excessively.
Autonomy	Unable to complete entire procedure, even in a straight forward case and with extensive verbal guidance.		Able to complete operation safely with moderate prompting.		Able to complete operation independently without prompting.
Level of difficulty	Easy exploration and dissection.		Moderate difficulty (e.g., mild inflammation, scarring, adhesions, obesity, severity of Disease.		Extremely difficult (e.g., severe inflammation, scarring, adhesion, obesity, or severity of disease).

CAT - Laparoscopic supracervical hysterectomy **ID:** **Assessor:** **Date:** **Case nr:**

Tasks	Instruments		Tissue		Errors		End-product	
Lig. mobilisation. Adequate exposure and dividing of the round ligament and the broad ligament.	**Use of instruments:**		**Use of non-dominant hand :**		**This task was performed with:**		**Was the ligaments divided safely?**	
	Uncoordinated	Stiff and uncontrolled movements,overshootin	Stagnant	NDH does not move	Complication	Bleeding	No	Divided in an unfavorable distance from the uterus.
	Hesitant	Controlled movements but hesitant and inefficient	Lagging	NDH is adjusting with delay or without efficiency	Near miss	To close to the pelvic wall. Bloody dissection.	Vaguely	No access to the vesicouterine space
	Skillful	Smooth, controlled and meaningful movements	Meaningful	Meaningful adjustment of NDH to improve exposure	No damage	No damage	Yes	Safe divition.
	Versatile	Masterful instrument use, effective movements	Forward looking	Strategic and intelligent adjustment by NDH	Tissue protective	Performed with best possible tissue protection	Anatomically	Crystal clear demonstration of anatomy
	n/a		n/a		n/a		n/a	
Adnex The proper ovarian ligament and fallopian tube are coagulated and divided	**Use of bipolar forceps, scissors or harmonic:**		**Use of non-dominant hand:**		**This task was performed with:**		**Was the proper ovarian ligament and the fallopian tube coagulated and divided correctly?**	
	Uncoordinatet	Stiff and uncontrolled movements, overshooting	Stagnant	NDH does not move	Complication	Bleeding	No	Divided in an unfavorable distance from the uterus.
	Hesitant	Controlled movements but hesitant and inefficient	Lagging	NDH is adjusting with delay or without efficiency	Near miss	To close to the ovarian tissue/uterus	Vaguely	To close to the ovarian tissue.
	Skillful	Smooth, controlled and meaningful movements	Meaningful	Adjustment of NDH to improve exposure	No damage	No damage to the ovaries.	Yes	Main structures identified and divided correctly.
	Versatile	Masterful instrument use, effective movements	Forward looking	Strategic and adjustment by NDH	Tissue protective	Performed with best possible tissue protection.	Anatomically	Crystal clear demonstration of anatomy
	n/a		n/a		n/a		n/a	

Vascular controle Identify the uterine vessels, coagulate and divide them	**Use of bipolar forceps and scissors:**		**Dissection of the uterine vessels:**		**This task was performed with:**		**Are the vessels identified and secured at the right level ?**	
	Hazardous	Uncontrolled movements	Hazardous	Insufficient view, uncontrolled movements	Complication	Bleeding, the vessels was not identified	Uncontrolled	Vessels not secured
	Laborious	Awkward and repeated unnecessary attempts	Laborious	Repeated unnecessary attempts	Near miss	Several attempts at several places to secure the vessels	Imprecise	Vessels not accurately secured
	Efficient	Instruments accurately placed and engaged	Efficient	Instruments accurately placed and engaged	Efficient	Visualisation of the vessels	Safe	Vessels secured before divition
	Masterly	Highly efficient and safe use of instruments	Masterly	Highly efficient and safe use of instruments	Ideal precision	Smooth and efficient dissection	Flawless	Perfectly secured before divition
	n/a		n/a		n/a		n/a	
Amputation Safe amputation	**Placement of the loop**		**Use of the non- dominant hand?**		**This task was performed with:**		**Was the uterus safely amputated ?**	
	Hazardous	Insufficient view, uncontrolled movements	Stagnant	NDH does not move	Complication	Intestines in the loop	Wrong plane	To close to the bladder, vagina
	Laborious	Awkward and repeated unnecessary attempts	Lagging	NDH is adjusting with delay or without efficiency	Near miss	No inspection of the position of the loop	Inconsistent	Below the secured vessels
	Efficient	Instruments accurately placed and engaged	Meaningful	Meaningful adjustment of NDH to improve exposure	Efficient	No damage	Satisfactory	Correct plane
	Masterly	Highly efficient and safe use of instruments	Forward looking	Strategic and intelligent adjustment by NDH	Ideal precision	Best possible tissue protection	Ideal plane	Correct plane, highly efficient
	n/a		n/a		n/a		n/a	

Teoretisk test i basal laparoskopi.
Udviklet af Jeanett Strandbygaard et al. i 2010 på Gynækologisk og Obstetrisk afdeling, Rigshospitalet, Københavns Universitets Hospital, til brug i specaillægeuddannelsen i gynækologi i Region Hovedestaden og Region Sjælland.
Spørgsmål bedes stilles til Jeanett Strandbygaard, læge, ph.d., på mail: jeanett78@gmail.com
Udviklingsprocessen er publiceret i Surgical Endoscopy 2013 27(4):1353-9

Test i basal laparoskopisk kirurgi for introduktionslæger i gyn.obs.

1) Ved en laparoskopisk operation bliver skærmen pludselig sort, efter at have virket normalt.
Hvilken af følgende er mest sandsynligt årsagen dertil?
a) Diatermiapparatet er slået fra
b) Sprunget lyspære
c) Skopet er dugget til
d) Insufflatorslangen er faldet af

B

2) Under en laparoskopisk operation, opstår der på skærmbilledet dårlig oversigt på grund af manglende distensering af abdomen.
Hvilken af følgende er mest sandsynligt årsagen dertil?
a) Defekt insufflator
b) Defekt optik
c) Skærmen
d) Lyskablerne

A

3) En slank, rask kvinde med BMI på 18 skal laparoskoperes, og hun kan være en risikopatient.
Hvilken af følgende er mest sandsynligt årsagen dertil?
a) Slanke patienter er svære at relaksere med anæstesi
b) Snævre arbejdsforhold i abdomen og pelvis
c) Risiko for lædering af store kar ved placering af Veress kanyle og primær trocar
d) Pneumoperitoneum er svært at opretholde hos slanke patienter

C

4) Du skal udføre en gynækologisk laparoskopi, og i den forbindelse orientere operationspersonalet om hvilken operationslejring, der er den mest anvendte ved gynækologisk laparoskopiske operationer.
Hvilken af følgende lejringer er mest oplagt under indgrebet?
a) Trendelenburgs leje
b) Fladt på ryggen
c) Anti-trendelenburg
d) 20° venstre sideleje

A

5) Du er i gang med en laparoskopisk operation på en patient der tidligere er opereret i abdomen, og du finder adhærencer spredt ud i abdomen og pelvis.
Hvilken af følgende vil være mest oplagt at foretage sig?

a) Bede anæstesien om at give mere relakserende
b) Sørge for at insufflere rigeligt med CO_2
c) Konvertere til laparotomi
d) Kalde på assistance fra en erfaren operatør

D

6) En erfaren operatør føler sig utryg ved at placere Veress kanyle i umbilicus pga. stor midtlinie incision, og må derfor finde et andet sted at anlægge pneumoperitoneum.
Hvilken af følgende steder er det mest sandsynligt, at operatøren så vil placere Veress kanyle?

a) I medioclaviculærlinien ud for costa 9 i højre side
b) I medioclaviculærlinien ud for costa 9 i venstre side
c) Caudalt for midtlinieinscisitionen
d) Man vil konvertere til laparotomi

B

7) Du er i gang med at placere Veress kanyle for at anlægge pneumoperitoneum på en patient, der er normalvægtig og ikke tidligere opereret i abdomen.
Hvilken af følgende er den mest oplagte måde at indføres Veress kanyle på?

a) Med lukket hane og med insufflatorslange påsat
b) Med åbenhane og med insufflatorslange påsat
c) Med lukket hane og uden insufflatorslange
d) Med åben hane og uden insuflatorslange

D

8) Du er i gang med at anlægge pneumoperitoneum, og har placeret Veress kanyle, men det intraabdominale tryk stiger ikke trods et stort insuffleret volumen.
Hvilken af følgende er mest sandsynligt årsagen dertil?

a) At kanylespidsen ligger i en tarm
b) Manglende relaksation af patienten
c) Utæthed i insufflationssystemet
d) Trykket der insuffleres med er for lavt

C

9) Du er ved at anlægge pneumoperitoneum, og har placeret Veress kanyle, men der er højt tryk under insufflationen.

Hvilken af følgende er ***mindst*** sandsynlig årsagen dertil?
a) Veress kanyle ligger ikke intraperitonealt
b) Patienten ikke er relakseret
c) Tynd patient
d) Veress kanyle er blokeret

C

10) Du er ved at insufflere CO_2, og vil gerne teste om du opnår pnuemoperitoneum.
Hvilken af følgende er mest oplagt?
a) Undersøge for ophævet dæmpning over leveren
b) Stetoskopere efter tarmlyde
c) Observere blodtryksstigning
d) Observere blodtryksfald

A

11) Under initial anlæggelse af pneumoperitoneum på en normalvægtig patient insuffleres en mængde CO_2.
Hvilken af følgende vil være den mest sandsynlige volumen?
a) 0.5 - 1 liter
b) 3 - 4 liter
c) 8 - 9 liter
d) 10-13 liter

B

12) Du er ved at indføre Veress kanyle og primær trocar via umbilicus på en normalvægtig patient, og er derfor opmærksom på anatomien.
Hvilken af følgende vil være mest oplagt at være opmærksom på?
a) Arteria epigastrica inferior
b) Arteria ovarica
c) Arteria iliaca communis
d) Arteria uterina

C

13) Du er ved at anlægge pneumoperitoneum, og finder lavt tryk intraabdominalt og intet gasflow.
Hvilken af følgende er mest sandsynligt årsagen?
a) Mangelfuld relaksation af patienten
b) Man befinder sig i et hulorgan
c) Insufflatoren er ikke aktiveret
d) Hul på et hulorgan

C

14) Du har anlagt pneumoperitoneum på en normalvægtig person, og har opnået et ønskværdig intraabdominalt tryk til din operaion.
Hvilket af følgende tryk er mest oplagt?

a) 1-5 mmHg
b) 10-15 mmHg
c) 20-25 mmHg
d) 30-35 mmHg

B

15) Du har anlagt pneumoperitoneum, og skal til at placere laterale trocar.
Hvilken af følgende placeringer er mest oplagt?

a) Lateralt for arteria epigastrica inferior
b) Ca. 2 cm medialt fra rectusbugen
c) Medialt for arteria epigastric inferior
d) Ca. 2 cm over symfysen

A

16) Du har anlagt pneumoperitoneum, og skal til at placere laterale trocar, og er derfor opmærksom på anatomien.
Hvilken af følgende vil være mest oplagt at være opmærksom på?

a) Arteria epigastrica inferior
b) Aorta
c) Arteria iliaca communis
d) Arteria ovarica

A

17) Du har placeret primære trocar, og skal identificere arteria epigastrica inferior.
Hvilken af følgende er mest oplagt til identifikation?

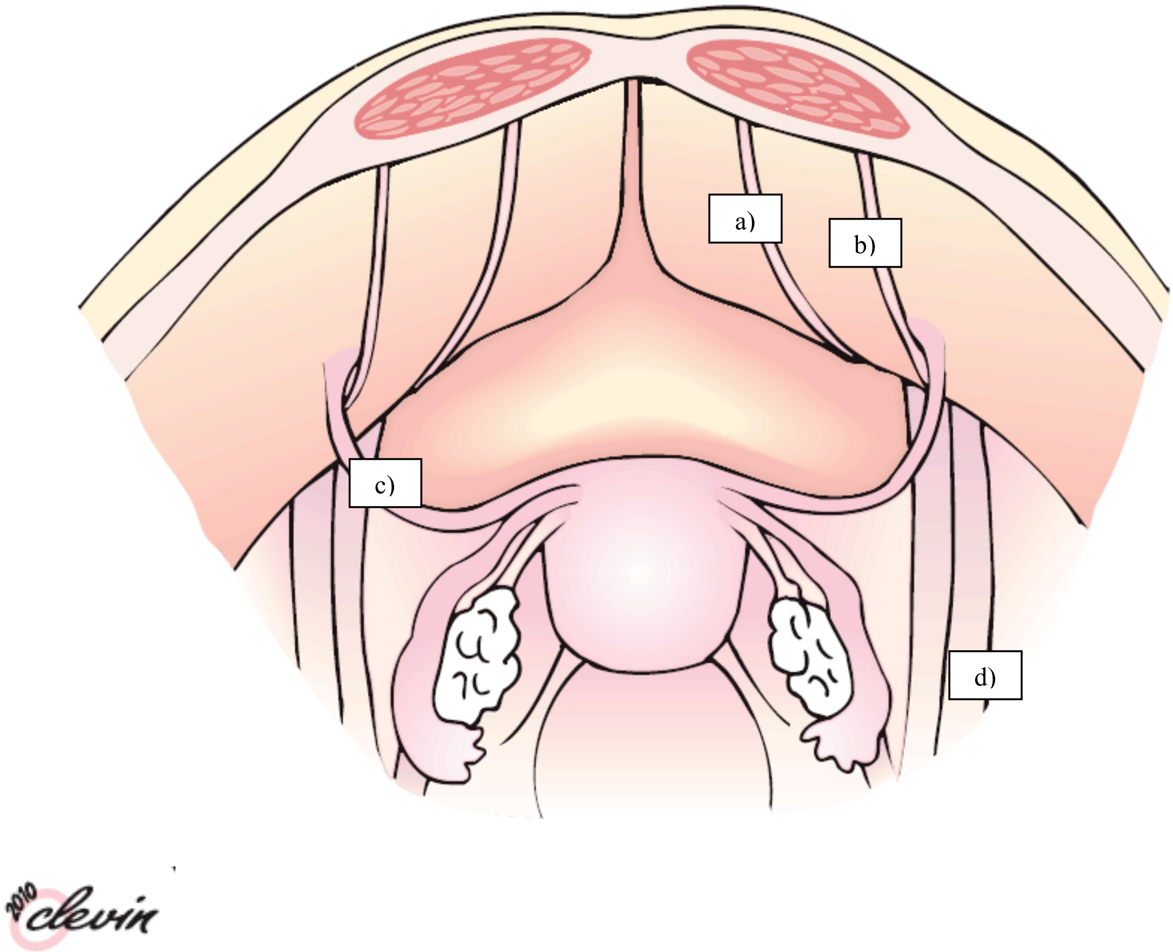
a) Gennemlysning af bugvæggen
b) Visuelt gennem optik i primær trocar
c) Palpation gennem bugvæggen
d) Identificering ved hjælp af ultralyd

B

Teoretisk test i basal laparoskopi.
Udviklet af Jeanett Strandbygaard et al. i 2010 på Gynækologisk og Obstetrisk afdeling, Rigshospitalet, Københavns Universitets Hospital, til brug i specaillægeuddannelsen i gynækologi i Region Hovedestaden og Region Sjælland.
Spørgsmål bedes stilles til Jeanett Strandbygaard, læge, ph.d., på mail: jeanett78@gmail.com
Udviklingsprocessen er publiceret i Surgical Endoscopy 2013 27(4):1353-9

18) Nævn de 4 strukturer markeret på tegningen.
Svar på svarark.

A) Obliterede arterie umbilicalis
B) Arteria epigastriga inferior
C) Ligamentum teres uteri (ligamentum rotundum)
D) Arteria iliaca externa

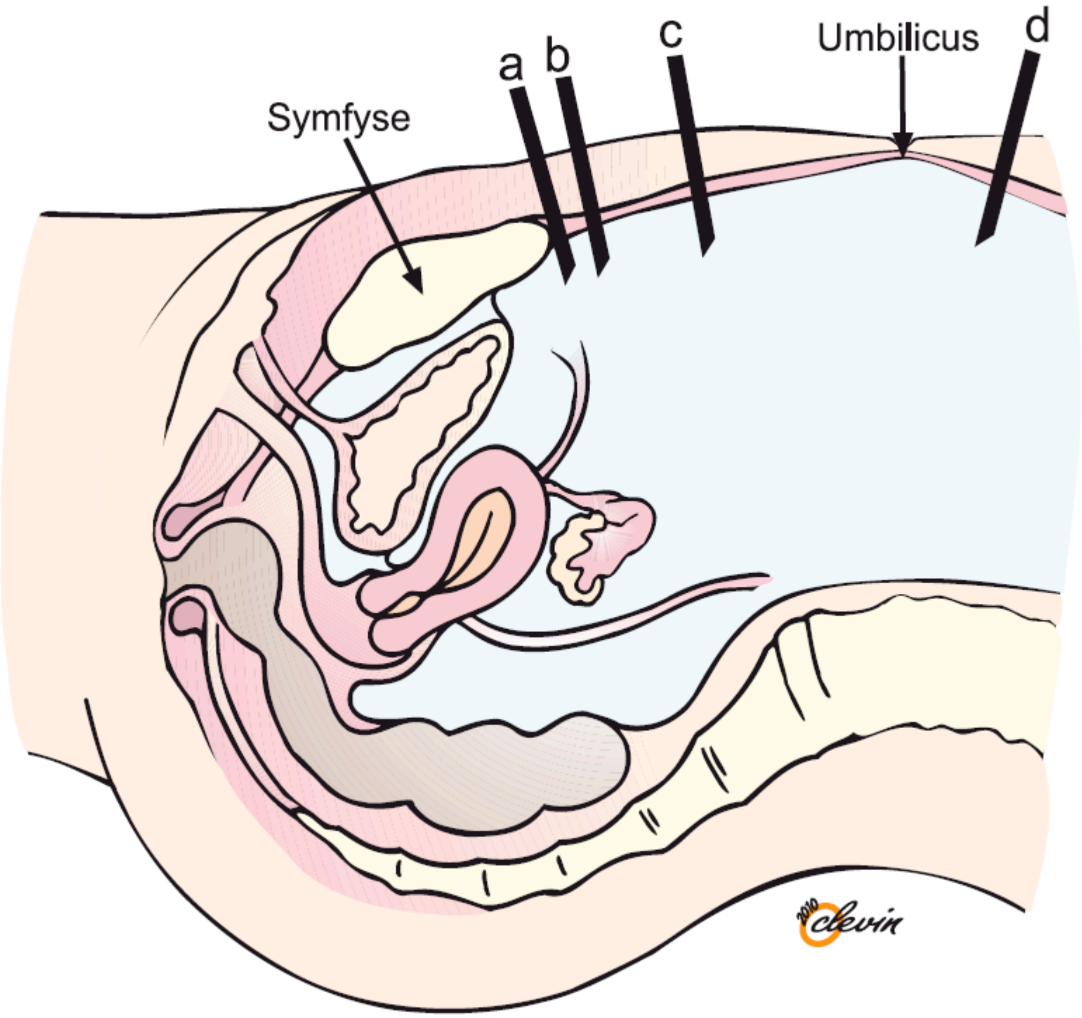


19) Du har ved placering af en trocar læderet arteria epigastrica inferior, og skal ud over at kalde hjælp, straks indlede behandling.
Hvilken af følgende er mest oplagt?

a) Fjerne trocaren
b) Yderlig insufflation af gas
c) Komprimering gennem trocaren med Foley kateter
d) Konvertere til laparotomi

C

20) Ved en laparoskopisk salpingectomi vil du gerne anlægge en suprapubisk adgang, og skal derfor placere en trocar.
Hvilken af følgende placeringer er mest oplagt?
Svar på svarark. B



21) Du er i gang med en gynækologisk laparoskopisk operation, og vil benytte dig af elkirugi?
Hvilken af følgende er mest oplagt?

a) Monopolær cut og coag
b) Monopolær coag
c) Bipolær cut
d) Bipolær coag

D

22) Du er i gang med en ukompliceret laparoskopisk operation, og anæstesipersonalet oplyser dig om, at patienten får en arytmi i form af en bradycardi.
Hvilken af følgende er mest oplagt initialt at gøre som gynækolog?

a) Insufflere et yderlig volumen gas
b) Fortsætte operationen uden tiltag
c) Bede om yderlig relaksation til patienten
d) Stoppe insufflation og lade gassen sive ud.

D

23) Du er i gang med en laparoskopisk operation hos en kvinde der tidligere har fået foretaget et sektio, og skal til at placere en suprapubisk trocar.
Hvilken af følgende er mest oplagt at overveje hos denne kvinde?

a) Blærens placering
b) Forløbet af ureter
c) Forløbet af arteria epigastrica inferior
d) Risiko for uterusruptur

A

24) Du er selvstændigt i gang med at operere en ekstrauterin graviditet laparoskopisk. Patienten er hæmodynamisk stabil, men du finder rigeligt med koagler og igangværende blødning som ikke er set på forudgående skanning. Du suger koagler med stort sug.
Hvilken af følgende er derudover mest oplagt at gøre?

a) Skynde sig at blive færdig med operationen
b) Scanne vaginalt for at se hvor meget blod der er i det lille bækken
c) Konvertere til laparotomi
d) Tilkalde assistance fra en mere erfaren operatør

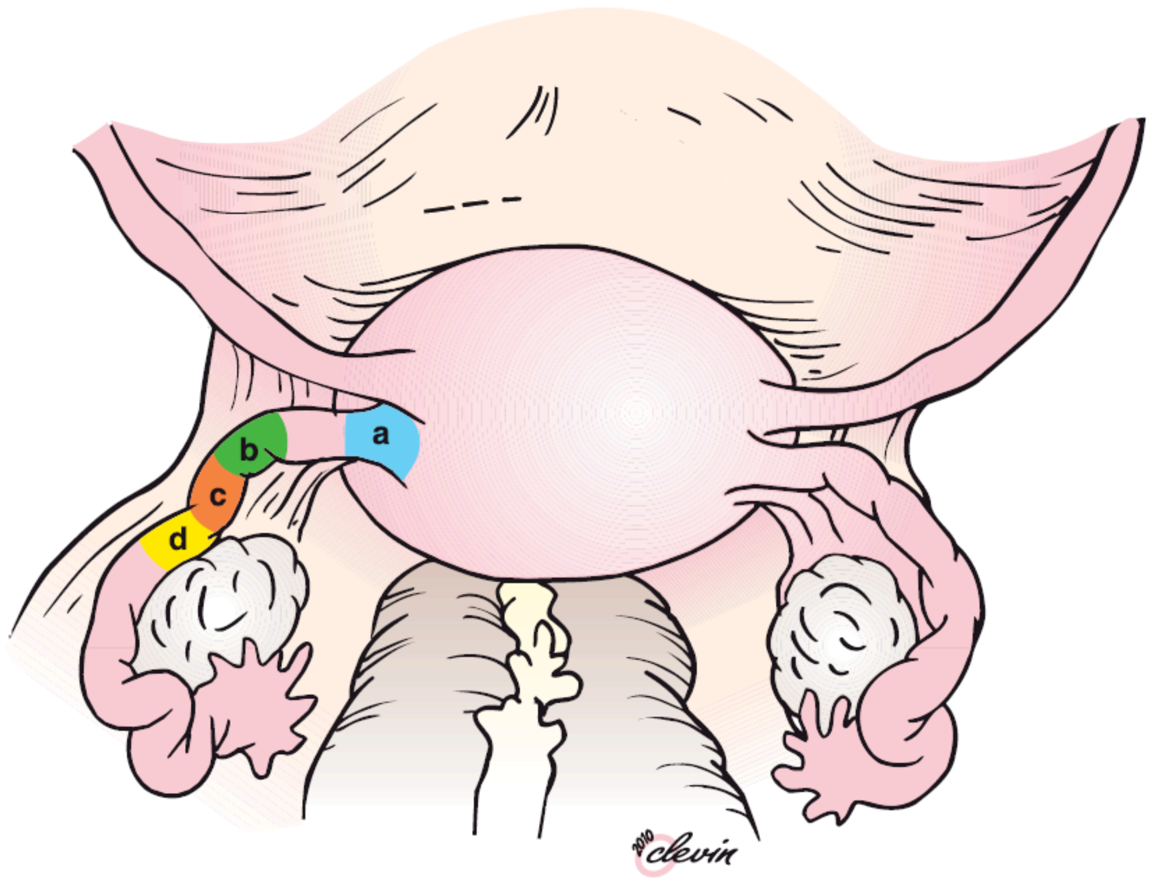
D

Teoretisk test i basal laparoskopi.
Udviklet af Jeanett Strandbygaard et al. i 2010 på Gynækologisk og Obstetrisk afdeling, Rigshospitalet, Københavns Universitets Hospital, til brug i specaillægeuddannelsen i gynækologi i Region Hovedestaden og Region Sjælland.
Spørgsmål bedes stilles til Jeanett Strandbygaard, læge, ph.d., på mail: jeanett78@gmail.com
Udviklingsprocessen er publiceret i Surgical Endoscopy 2013 27(4):1353-9

25) Du er i gang med en laparoskopisk salpingektomi, og skal til at brænde ved den laterale del af salpinx, og er derfor opmærksom på anatomien.
Hvilken af følgende vil være mest oplagt at være opmærksom på?
a) Arteria ovarica
b) Arteria uterina
c) Arteria epigastrica inferior
d) Arteria obturatorius

A

26) Du er i gang med en laparoskopisk salpingektomi, og skal til at fjerne salpinx ved uterinhjørnet.
Hvilken af følgende er mest sandsynlig det korrekte sted at fjerne salpinx?
Svar på svarark. A



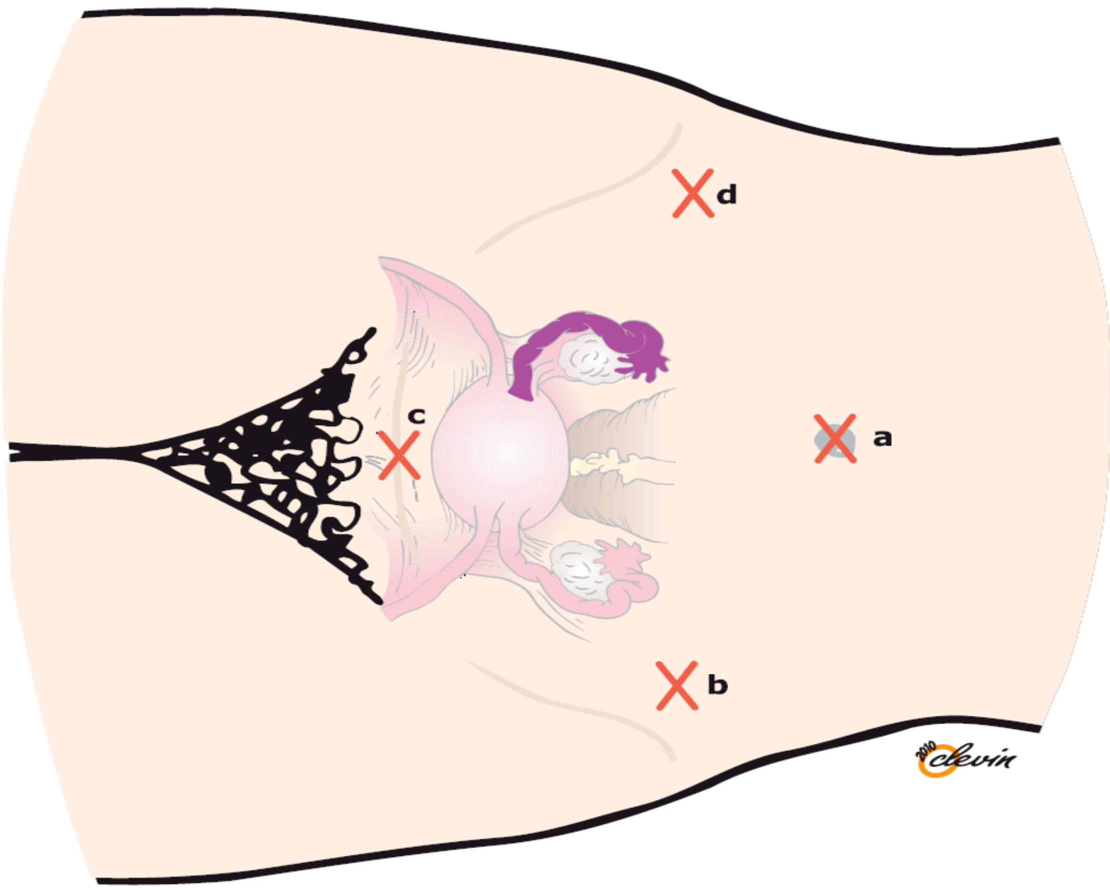
Teoretisk test i basal laparoskopi.
Udviklet af Jeanett Strandbygaard et al. i 2010 på Gynækologisk og Obstetrisk afdeling, Rigshospitalet, Københavns Universitets Hospital, til brug i specaillægeuddannelsen i gynækologi i Region Hovedestaden og Region Sjælland.
Spørgsmål bedes stilles til Jeanett Strandbygaard, læge, ph.d., på mail: jeanett78@gmail.com
Udviklingsprocessen er publiceret i Surgical Endoscopy 2013 27(4):1353-9

27) Under en selvstændig laparoskopisk operation er du usikker på, om du kommer til at lave en termisk skade på tarmen, og derfor tilkalder du en mere erfaren kollega.
Mens du venter er følgende desuden oplagt at gøre?

a) Kalde mavetarmkirurgisk bagvagt med det samme
b) Konvertere til laparotomi
c) Begynde 3 stofs antibiotikabehandling
d) Relaksere patienten yderligere

C

28) Du er i gang med en laparoskopisk operation sammen med en erfaren kollega. I bliver nød til at supplere med en mini-laparotomi, og må derfor udvide en trocaradgang.
Hvilken af følgende adgange vil være mest oplagt at udvide?
Svar på svarark. C



29) Efter et afsluttet laparoskopisk indgreb lejres patienten med henblik på optimalt af få eksuffleret gas.
Hvilken af følgende lejringer er mest oplagt?

a) Trendelenburg
b) Antitrendelenburg
c) 20° højre sideleje
d) 20° venstre sideleje

A

30) Du er i gang med eksufflation, og det er vigtigt fortsat at beholde trocars i med ventilen åben, til gassen er ude.
Hvilken af følgende er mest oplagt som begrundelse for dette?
a) Ellers kan gassen ikke komme ud
b) Man undgår at få tarm eller en omentsnip med ud
c) Man eliminerer postoperative skuldersmerter
d) Man reducerer postoperativ infektionsrisiko

B

31) Under en laparoskopisk operation undersøges altid for eventuelle blødninger, men det er særlig vigtig, at undersøge for venøs blødning på et bestemt tidspunkt.
Hvilken af følgende muligheder er mest sandsynlig?
a) Ved initial anlæggelse af pneumoperitoneum
b) Ved placering af lateralt trocar
c) Ved placering af suprapubisk trocar
d) Ved eksufflation

D

32) Efter fjernelse af trocars ses en en fasciedefekt på over 12 mm, og den skal sutureres.
Hvilken af følgende er mest sandsynlig årsagen dertil?
a) For at forhindre infektion
b) For at undgå herniering
c) For at forhindre ardannelse
d) For at undgå blødning

B

33) Du har ved en laparoskopisk operation fjernet salpinx hos en patient på grund af en tubar graviditet, og operationen forløb ukompliceret.
Hvilken af følgende vil det være mest oplagt at foretage postoperativt?
a) Scanne efter 7 dage
b) Scanne efter 14 dage
c) Følge s-hCG 7 dage postoperativt

d) Udskrive patienten uden yderlig kontrol

D

34) En patient der er blevet opereret laparoskopisk for endometriose bliver på 5. –6. postopeative dag dårlig med feber, ondt i maven og har ændret afføring.
Hvilken af følgende er mest sandsynlige årsagen dertil?

a) Intraabdominal blødning
b) Ureterlæsion
c) Termisk tarmlæsion
d) Nosokomiel gastrointestinal infektion

C

LSH questions

1. Du skal operere en pasient som skal til laparoskopisk supracervikal hysterektomi (LSH). Hva vil du informere henne om før operasjonen? Nevn de tre viktigste momenter.
2. Hvilke prøver vil du ta av en pasient som skal til en LSH? (sett et kryss)

☐ Hb ellers er ingen spesielle prøver nødvendige
☐ Hb og cervix cytologisk prøve ellers er ingen prøver nødvendige
☐ Hb, cervix cytologisk prøve og en endometrie biopsi (pipelle)

3.Hvilket tiltak mtp port plassering kan bedre tilgangen ved en stor uterus.

4. Du skal gjøre en LSH og skal eksponere istmus/cervix og skyver blæren ned. Du er usikker på om du har fått en blæreskade. Hva gjør du?

☐ Instillerer blåfarge i blæren via et transuretralt kateter og ser etter lekkasje.

☐ Fortsetter operasjonen og bestiller en urografi til neste dag.

☐ Følger med på kreatinin verdiene de første postoperative dager.

5. Ureter skades lettest på to steder under en LSH +BSOE
Marker disse to områdene på tegningen og hvilken aktuell gynekologisk struktur som ligger i nærheten.

Tegning:

I

Assessment of Surgical Competence: Development and Validation of Rating Scales Used for Laparoscopic Supracervical Hysterectomy

Jeanne Mette Goderstad, MD,[*] Leiv Sandvik, MSc, PhD,[†] Erik Fosse, MD, PhD,[‡,§] and Marit Lieng, MD, PhD[§,‖]

[*]Department of Gynecology, Sørlandet Hospital, Arendal, Norway; [†]Oslo Center for biostatistics and epidemiology, Oslo University Hospital, Oslo, Norway; [‡]The Intervention Centre, Oslo University Hospital, Oslo, Norway; [§]Department of Gynecology, Oslo University Hospital, Oslo, Norway; and [‖]Institute of Clinical Medicine, University of Oslo, Oslo, Norway

OBJECTIVE: To develop a procedure-specific rating scale for laparoscopic supracervical hysterectomy (LSH), and to compare the construct validity and reliability with a general rating scale in laparoscopic surgery, global operative assessment of laparoscopic skills (GOALS).

DESIGN: Prospective interobserver study. In collaboration with an expert group, we developed the procedure-specific rating scale, competence assessment tool for laparoscopic supracervical hysterectomy (CAT-LSH). LSH was performed by gynecologists with different levels of surgical competence levels (13 procedures were performed by inexperienced trainees, 13 by intermediate experienced, and 15 by laparoscopic experts). All procedures were video-recorded. Surgical performance was evaluated in all procedures using both CAT-LSH and GOALS by the surgical assistant, as well as by 2 blinded observers evaluating the video recordings.

SETTING: University teaching hospital.

PARTICIPANTS: Laparoscopic experts, consultants and gynecological registrars from the Department of Gynecology.

RESULTS: There were significant differences between the 3 proficiency groups in both the rating scales. Mean GOALS score evaluated by the operating assistant and the 2 observers were for inexperienced surgeons 16.4 vs. 13.6 ($p < 0.01$), for surgeons with intermediate experienced 22.6 vs. 19.5 ($p < 0.05$) and for expert surgeons 26.1 vs. 22.4 ($p < 0.01$), respectively. Corresponding results for the CAT-LSH scores were 41.0/34.6 ($p < 0.01$), 49.2/43.1 ($p < 0.01$), and 58.7/51.1 ($p < 0.01$), respectively. The interrater reliability measured by the interclass correlation coefficient between the surgical assistant and the 2 blinded observers for GOALS and CAT-LSH were 0.71 and 0.75, respectively.

CONCLUSIONS: The GOALS and CAT-LSH appear to have construct validity and high interrater reliability. Assessment of surgical competence during LSH is feasible in daily practice with objective rating scales like CAT-LSH and GOALS. (J Surg Ed 73:600-608. © 2016 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: surgical skills, surgical competence, laparoscopic hysterectomy, GOALS

COMPETENCIES: Medical Knowledge, Interpersonal and Communication Skills, Professionalism, Practice-Based Learning and Improvement

INTRODUCTION

The ultimate goal for a surgical training program is to produce competent clinical professionals capable of meeting the healthcare needs of the society.[1] Clinical competence is achieved by development of a combination of cognitive factors, ehavioral, perceptual, and technical skills.[2,3] Competencies can be learned, and feedback is decisive to achieve good learning.[4] Feedback works at its best when it inspires to reflection on behavior and performance.[5] Medical

A scholarship covered for the expenses related to the memory sticks for the video recordings. The Department of Gynecology, Oslo University Hospital, covered the rest of the expenses of the study.

Correspondence: Inquiries to Jeanne Mette Goderstad, MD, Department of Gynecology, Sørlandet Hospital, Arendal, Sykehusveien, 4838 Arendal, Norway; fax number: +4737014041; E-mail: jeanne.mette.goderstad@sshf.no, jmgoderstad@hotmail.com

1931-7204/$30.00
http://dx.doi.org/10.1016/j.jsurg.2016.01.001

trainees want feedback from their supervisors and superiors, but they are often disappointed by the small amount of feedback and how it is provided.[6] Trainees need systematization of feedback where criteria for good performance are defined in advance and evaluation is part of the training program.[7] The ideal assessment tool produces reliable, valid results and is furthermore practical.[1]

Technical skills are essential in surgical performance. Several tools for assessment of technical skills in minimal invasive procedures have been developed.[8] These include expert-based evaluation, proficiency-based tools, objective motion analysis, and virtual reality simulators.[9] The objective of rating scales is the formalization of the traditional model of assessment of surgical trainees by more experienced surgeons. This is achieved by standardization and objectification of subjective data.[10,11] Medical education is undergoing a shift from the experience-based model to documentation of proficiency.[12,13] In competency-based training, objective assessment of performance is an essential component. Standardized assessment tools may be general or procedure specific. The general assessment tool, global operative assessment of laparoscopic skills (GOALS), has previously been validated for laparoscopic cholecystectomy, ventral hernia repair, and appendectomy.[8,11,14] Individual factors influence the technical difficulty experienced in performing laparoscopic hysterectomy. These include patient obesity; adhesions from previous surgery or infections; and the size, location, and number of uterine fibroids. To try to equalize the laparoscopic supracervical hysterectomy (LSH) procedures by correction for casemix, we chose the traditional GOALS, expanded with 1 item, degree of difficulty (Table 2).[11,14] In this study, we aimed to develop a procedure-specific assessment tool for LSH. We called it competence assessment tool for laparoscopic supracervical hysterectomy (CAT-LSH) (Table 1). CAT-LSH was developed as a modification of previous described and validated ratings scales for different surgical procedures, developed at Imperial College in London, UK.[15] Furthermore, we aimed to compare the construct validity and reliability of a general rating scale GOALS (Table 2) and the procedure-specific rating scale (CAT-LSH) on the performance of a LSH.

METHODS

This is a prospective interobserver study. To develop the CAT-LSH we conducted interviews with international experts in gynecological laparoscopy with recognition within formal surgical education. The experts used in our study are all recognized for their proficiency in endoscopic surgery by the European Society of Gynecological Endoscopy and American Association of Gynecologic Laparoscopists. Total 5 experts were contacted by e-mail, 3 of which kindly agreed to contribute to our study. Our interviews with the experts focused on development of a surgical strategy of LSH, defining the main steps of the procedure, the surgical challenges for each step of the procedure, and the structure of the assessment sheet. The assessment sheet and the procedure strategy were developed from the results of our interviews, and were finalized following review by the contributing experts. We identified 4 steps of the surgical strategy—ligament mobilization, release of the adnexa from the uterus, division of the uterine vessels, and uterus amputation. We chose to omit retrieval of tissue by morcellation because of the difference in size and consistence of the myoma, the difference in diameter of different morcellators and the skills of the assistant. For each step use of instruments, tissue handling, errors, and the end product are evaluated. The surgeon gets a score from 1 to 4 for each step, which gives a maximum score of 16 points for each step of the procedure and a maximum total score of 64 points for the entire procedure (Table 1).

We included specialists and specialist trainees with different proficiency levels currently working at the Department of Gynecology, at our teaching university hospital. All study participants signed an informed consent at time for inclusion. The surgical experience of each study participants was registered and they were categorized as being inexperienced (previously never performed laparoscopic hysterectomies and less than 50 laparoscopic procedures in total), with intermediate experience (previously performed more than 5 LSH procedures and more than 50 laparoscopic procedures in total), and experts (senior consultants at the endoscopic unit who perform advanced pelvic surgery). This categorization of experience was based on a previous study on laparoscopic training.[16]

All study participants performed 1 to 3 LSH procedures and the procedures were video-recorded and stored on individual memory sticks. The surgical performance was evaluated using GOALS and CAT-LSH by the surgical assistant as well as 2 blinded observers evaluating the surgical procedure on video footage. All LSH surgical procedures were performed with bipolar desiccation, grasper and scissors. Other, alternative surgical tools were not used to standardize the procedure.

The operating assistant (specialist trainee or consultant) received information about the rating scales before the surgical procedure. Immediately after surgery, the operating assistant filled in the assessment forms. We used the operating assistant planned for the surgery, and did not make any changes to the planned surgical schedule because of the study. The 2 observers who evaluated the video recordings are experts in laparoscopic surgery and work at the department. They were blinded as to the identity of the study participants and the proficiency level. The memory sticks with the taped surgical procedure were marked with a number and put in separate envelopes. Each envelope contained 1 memory stick and 1 blank copy of GOALS

TABLE 1. Competence Assessment Tool for Laparoscopic Supracervical Hysterectomy—CAT-LSH

CAT-Laparoscopic			Supracervical	Hysterectomy	ID	Assessor	Date	Case Number
Tasks	**Instruments**		**Tissue**		**Errors**		**End product**	
Ligament Mobilization	**Use of Instruments**		**Use of Nondominant Hand**		**This Task Was Performed With**		**Was the Ligaments Divided Safely?**	
Adequate exposure and dividing of the round ligament and the broad ligament.	Uncoordinated 1p	Stiff and uncontrolled movements, overshootin	Stagnant 1p	NDH does not move	Complication 1p	Bleeding	No 1p	Divided in an unfavorable distance from the uterus.
	Hesitant 2p	Controlled movements but hesitant and inefficient	Lagging 2p	NDH is adjusting with delay or without efficiency	Near miss 2p	To close to the pelvic wall. Bloody dissection.	Vaguely 2p	No access to the vesicouterine space
	Skillful 3p	Smooth, controlled and meaningful movements	Meaningful 3p	Meaningful adjustment of NDH to improve exposure	No damage 3p	No damage	Yes 3p	Safe division.
	Versatile 4p	Masterful instrument use, effective movements	Forward looking 4p	Strategic and intelligent adjustment by NDH	Tissue protective 4p	Performed with best possible tissue protection	Anatomically 4p	Crystal clear demonstration of anatomy
	n/a		n/a		n/a		n/a	sum
Adnex	**Use of Bipolar Forceps, Scissors or Harmonic**		**Use of Nondominant Hand**		**This Task Was Performed With**		**Was the Proper Ovarian Ligament and the Fallopian Tube Coagulated and Divided Correctly?**	
The proper ovarian ligament and fallopian tube are coagulated and divided	Uncoordinated 1p	Stiff and uncontrolled movements, overshooting	Stagnant 1p	NDH does not move	Complication 1p	Bleeding	No 1p	Divided in an unfavorable distance from the uterus.
	Hesitant 2p	Controlled movements but hesitant and inefficient	Lagging 2p	NDH is adjusting with delay or without efficiency	Near miss 2p	To close to the ovarian tissue/ uterus	Vaguely 2p	To close to the ovarian tissue.
	Skillful 3p	Smooth, controlled and meaningful movements	Meaningful 3p	Adjustment of NDH to improve exposure	No damage 3p	No damage to the ovaries.	Yes 3p	Main structures identified and divided correctly.
	Versatile 4p	Masterful instrument use, effective movements	Forward looking 4p	Strategic and adjustment by NDH	Tissue protective 4p	Performed with best possible tissue protection.	Anatomically 4p	Crystal clear demonstration of anatomy
	n/a		n/a		n/a		n/a	sum

Vascular Control	Use of Bipolar Forceps and Scissors		Dissection of the Uterine Vessels		This Task Was Performed With		Are the Vessels Identified and Secured at the Right Level?	
Identify the uterine vessels, coagulate and divide them	Hazardous 1p	Uncontrolled movements	Hazardous 1p	Insufficient view, uncontrolled movements	Complication 1p	Bleeding, the vessels was not identified	Uncontrolled 1p	Vessels not secured
	Laborious 2p	Awkward and repeated unnecessary attempts	Laborious 2p	Repeated unnecessary attempts	Near miss 2p	Several attempts at several places to secure the vessels	Imprecise 2p	Vessels not accurately secured
	Efficient 3p	Instruments accurately placed and engaged	Efficient 3p	Instruments accurately placed and engaged	Efficient 3p	Visualization of the vessels	Safe 3p	Vessels secured before division
	Masterly 4p	Highly efficient and safe use of instruments	Masterly 4p	Highly efficient and safe use of instruments	Ideal precision 4p	Smooth and efficient dissection	Flawless 4p	Perfectly secured before division
	n/a		n/a		n/a		n/a	sum

Amputation	Placement of the Loop		Use of the Nondominant Hand?		This Task Was Performed With		Was the Uterus Safely Amputated?	
Safe amputation	Hazardous 1p	Insufficient view, uncontrolled movements	Stagnant 1p	NDH does not move	Complication 1p	Intestines in the loop	Wrong plane 1p	To close to the bladder, vagina
	Laborious 2p	Awkward and repeated unnecessary attempts	Lagging 2p	NDH is adjusting with delay or without efficiency	Near miss 2p	No inspection of the position of the loop	Inconsistent 2p	Below the secured vessels
	Efficient 3p	Instruments accurately placed and engaged	Meaningful 3p	Meaningful adjustment of NDH to improve exposure	Efficient 3p	No damage	Satisfactory 3p	Correct plane
	Masterly 4p	Highly efficient and safe use of instruments	Forward looking 4p	Strategic and intelligent adjustment by NDH	Ideal precision 4p	Best possible tissue protection	Ideal plane 4p	Correct plane, highly efficient
	n/a		n/a		n/a		n/a	sum

NDH, nondominant Hand.

TABLE 2. GOALS—Global Operative Assessment of Laparoscopic Skills

Domains	1	2	3	4	5
Depth perception	Constantly overshooting target, hits backstop, wide swings, slow to correct.		Some overshooting or missing plane but corrects quickly.		Accurately directs instruments in correct plane to target.
Bimanual dexterity	Use of 1 hand, ignoring nondominant hand, poor coordination between hands.		Use of both hands but does not optimize interactions between hands to facilitate conduct of operation.		Expertly uses both the hands in a complementary manner to provide optimal working exposure.
Efficiency	Uncertain, much wasted effort, many tentative motions, constantly changing focus of operation, or persisting at a task without progress.		Slow, but planned and reasonably organized.		Confident, efficient, and safe conduct of operation, maintaining focus on component of procedure until better done approach.
Tissue handling	Rough, tears tissue by excessive traction, injures adjacent structures, poor control of coagulation device, grasper frequently slips off.		Handles tissues reasonably well with some minor trauma to adjacent tissues, for example, coagulation of nontarget tissue, occasional slipping of grasper.		Handles tissue very well with appropriate traction on tissues and negligible injury of adjacent structures. Uses energy sources appropriately but not excessively.
Autonomy	Unable to complete entire procedure, even in a straight forward case and with extensive verbal guidance.		Able to complete operation safely with moderate prompting.		Able to complete operation independently without prompting.
Level of difficulty	Easy exploration and dissection.		Moderate difficulty (e.g., mild inflammation, scarring, adhesions, obesity, severity of disease).		Extremely difficult (e.g., severe inflammation, scarring, adhesion, obesity, or severity of disease).

and CAT-LSH marked with the same number. The 2 observers received identical envelopes. The study participants did not receive any information about their performance scores.

Statistics

A pilot study was conducted to calculate the sample size of the study. A total of doctors at the department with different surgical experience (same categories as in the main study) were included in the pilot study. They performed a LSH and the performance was evaluated by the surgical assistant and the blinded observers using GOALS and CAT-LSH. Based on the results of the pilot study, we included 37 LSH procedures in the study to achieve a study power of 80 %, with at a level of significance of 0.05. Normally distributed continuous data from 2 study groups were compared using a 2-sided Independent Samples *t* test. To measure agreement between the assessors, we used the interclass correlation coefficient (ICC). Analysis were performed using the statistical Package for Social Science (SPSS) for Windows (SPSS Inc., Chicago, IL). Inclusion of participants started seventh of January 2013 and was completed by first of September the same year. The study

TABLE 3. Discriminative Validity Based on Proficiency Group for the 2 Rating Scales: Global Operative Assessment of Laparoscopic Skills (GOALS) and Competence Assessment Tool for Laparoscopic Supracervical Hysterectomy (CAT-LSH)

	Proficiency Group	GOALS (SD)	p Value	CAT-LSH (SD)	p Value
Observed by assistant surgeon (mean)	Inexperienced ($n = 11$)	16.4 (3.2)	<0.001	41.0 (5.0)	<0.001
	Intermediate experienced ($n = 12$)	22.6 (2.5)	0.002	49.2 (4.3)	<0.001
	Experts ($n = 14$)	26.1 (2.5)		58.7 (6.1)	
Blinded observers (mean)	Inexperienced ($n = 11$)	13.6 (3.3)	0.001	34.6 (4.3)	0.006
	Intermediate experienced ($n = 12$)	19.5 (4.1)	0.085	43.1 (8.2)	0.011
	Experts ($n = 14$)	22.4 (4.3)		51.1 (6.6)	

SD, standard deviation.

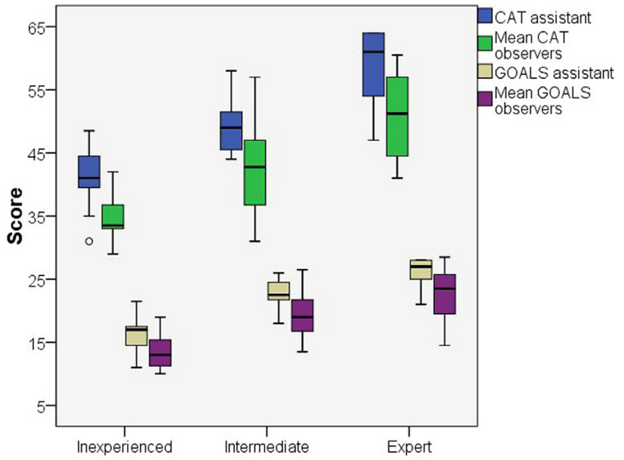


FIGURE. Box plot of the total score for the rating scales by the different assessors. GOALS: lowest possible score 6 and highest possible score 30. CAT-LSH: lowest possible score16 and highest possible score 64.

was approved by the head of the Department and the head of the Research Center for Obstetrics and Gynecology at our institution. Ethical approval from the Regional Ethical Committee was not needed.

RESULTS

All doctors invited to participate in the study, accepted the invitation. A total of 41 procedures were performed for witch 4 video recordings were incomplete and therefore discarded. Consequently, 37 procedures were assessed by the assistant surgeons and 2 blinded observers using both GOALS and CAT-LSH. In all procedures, 11 were performed by a total of 9 participants with lowest rated proficiency, 12 were performed by a total of 6 participants with intermediate proficiency, and 14 were performed by 6 participants with the highest rated proficiency. All the patients who were operated had proved consent for the video recording of the procedures. The indications for surgery were uterine fibroids and abnormal uterine bleeding. The median weight of the surgical specimen (morcellated uterine corpus) was 261 g (minimum 40 g and maximum 820 g). The degree of difficulty of the surgical procedure was evaluated with GOALS. There were no significant differences in this score between the 3 proficiency groups.

Discriminative Validity

The results of the surgical assessments by GOALS and CAT-LSH performed by the surgical assistants and the blinded observers are presented in Table 3. When performed by the assistant surgeon, assessment by GOALS differed significantly between inexperienced and intermediate experienced surgeons ($p < 0.001$), and intermediate surgeons and experts ($p = 0.002$). Similarly, assessment by GOALS by the blinded observers differed significantly between the inexperienced and intermediate experienced surgeons ($p = 0.001$). When intermediate experienced surgeons were compared to experts by the blinded observers using GOALS, the difference was not statistically significant ($p = 0.085$). When performed by the assistant surgeon, assessment by CAT-LSH also discriminated between inexperienced and intermediate experienced surgeons ($p < 0.001$) and intermediate experienced surgeons and experts ($p < 0.001$). Assessment by the blinded observers using CAT-LSH also differed significantly between the proficiency groups (inexperienced vs. intermediate, and intermediate vs. experts, $p = 0.006$ and $p = 0.011$, respectively). Both the total GOAL and CAT-LSH scores were significantly higher when scored by the assistant surgeon compared to when scored by blinded observers in all 3 proficiency groups (Table 3). The box plot of GOALS and CAT-LSH for each proficiency is demonstrated in the Figure.

Reliability

The reliability values comparing the mean scores for both GOALS and CAT-LSH of the 2 blinded observers assessing identical recordings, and between surgical assistants and the mean of the 2 blinded observers are presented in Table 4. Overall, the ICC of both GOALS and LSH-CAT indicated very good agreement when the 2 blinded observers were compared. As expected, the ICC declined when the mean score of the 2 blinded observers and the surgical assistants were compared.

DISCUSSION

Both the rating scales used in our study demonstrated construct validity. The interrater reliability was high. Evaluation with GOALS and CAT-LSH is feasible in daily

TABLE 4. Interrater reliability, ICC (0-1), for the rating scales; Global Operative Assessment of Laparoscopic Skills (GOALS) and Competence Assessment Tool for Laparoscopic Supracervical Hysterectomy (CAT-LSH)

	n	GOALS Observer 1/ Observer 2	GOALS Assistant Surgeon/Observers	CAT-LSH Observer 1/ Observer 2	CAT-LSH Assistant Surgeon/Obserevers
Total	37	0.74	0.71	0.85	0.75

Observer 1 and observer 2 are the blinded observers. Operating assistant are the different doctors assisting on the operation.

practice. The results of our study are consistent with results from previous studies evaluating both gynecological and surgical procedures.[11,14,16,17] In our study the assessment of the surgical procedure was done by the operating assistant and blinded observers. The operating assistant gave a higher total score, both by GOALS and CAT-LSH, in all the 3 proficiency groups compared to the blinded observers. The reason for this difference might be that the operating assistant knew the person who was performing the surgery and therefore tended to be more positive in the evaluation. One consequence of this difference is related to benchmarking for proficiency group in surgical procedures. If summative assessment is performed, assessment scores provided by operating assistants and by blinded video-assessment cannot be compared. Consequently, repeated assessment of surgical performance to evaluate a learning curve for a candidate should be performed either by the operating assistant or a blinded observer evaluating a video-recorded procedure. Ideally, the range of performance should be narrow within each proficiency group, indicating that the subjects fit the group definition. In our study, we found a narrow performance range in the inexperienced group and a wider performance range in the expert group. The results could be coincidental because of small sample size. It is difficult to standardize surgical procedures and another explanation for the wider performance range in the expert group might be that the experts treat the most complicated cases. The mean degree of difficulty score on GOALS was the same for the different proficiency groups in our study. Consequently, our results indicate that some gynecologists in the expert group are more skilled than others. Different competence levels within the expert group might also explain the small differences between the intermediate experienced surgeons and experts when the 2 blinded observers assessed the video-recorded procedures using GOALS. CAT-LSH appeared to have discriminative validity for all 3 proficiency groups. This may be because this form is more detailed regarding the performance at each step of the procedure compared to GOALS. Ahmed et al.[18] conclude in their review that a combination of global and task-specific assessment tools seems to be the most comprehensive solution for observational assessment tools of technical skills. The results of our study support this statement. The level of agreement between the different assessors is important in the evaluation of an assessment system. It reveals how unambiguous the test is, and thereby how valid it is if used by different independent raters.[19] A disadvantage of ICC is that it only gives a general value and does not make distinction between various types of disagreement. An ICC greater than 0.75 is considered good agreement.[19] In our study, ICC varied from 0.75 to 0.84. As expected, the lowest score was between the operating assistant and the blinded observers. This was the case for both GOALS and CAT-LSH. Overall, the ICC are acceptable for the 2 different assessment tools. Our results are consistent with results from similar studies.[11,16] The advantages of GOALS and CAT-LSH differ. GOALS is general and can thereby be used for evaluation of surgical skills at different laparoscopic procedures. This is a benefit as both the assessors and those who are being assessed can get experience with 1 assessment tool for different laparoscopic procedures. Another advantage of GOALS is the possibility to include the degree of difficulty of the procedure into the assessment. The technical difficulty during laparoscopic hysterectomy varies essentially with size, location, and number of fibroids. In our study, the weight of the smallest uterus was 40 g and the largest 820 g. With GOALS, there is a possibility to correct for this casemix. Assessment by CAT-LSH might be more time consuming as the assessment provides more detailed information about the performance during the different steps of the hysterectomy. As CAT-LSH gives an opportunity to evaluate each step of the hysterectomy, it might help the trainee to focus on the part of the procedure that needs improvement, and at the same time receive information about dexterity. The blinded observers in our study are experienced laparoscopic surgeons. Experts are not necessary for a good evaluation. Oestegaard et al.[20] have shown that video evaluation can be done by inexperienced doctors. We chose experts as observers because they were available and it was easier to use them than to train inexperienced persons. Based on the results of this study, we envisage future inclusion of GOALS and CAT-LSH as part of the training of young doctors in the operating theater. Intraoperative assessment can be completed at the end of each surgical procedure. This is less time consuming than the video evaluation. In some institutions, GOALS are computerized and automatically appears at the end of the operation's documentation.[14] An advantage with the video recording is that the surgeon has the possibility to watch the video and evaluate his own performance after the procedure. Both GOALS and LSH-CAT can be used for summative and formative assessment with the purpose to evaluate the test-takers abilities, which can be used for accreditation or to set a score that must be passed to receive a certification for laparoscopic hysterectomy.

A strength of our study is that we demonstrate that evaluation with GOALS and CAT-LSH is feasible in daily practice. Assessment by GOALS and CAT-LSH were performed on real surgical procedures. We were assessing surgical performance in the operating room, and not in an artificial model demonstrating how the study participants intend to perform surgery. We did not adjust the surgical schedule or the surgical assistant because of the study. The evaluation of the rating scales were performed during daily practice, which indicates that they may be used for assessment and evaluation of laparoscopic trainees during their regular work. The main limitation of the study is the relatively small sample size. Furthermore, a preinclusion test of the study participants would have strengthened the study by providing more similar surgical competence within the different proficiency groups. A self-evaluation of the surgical procedures by the surgeons themselves could also have added interesting information.[21]

Owing to the safety concerns regarding power morcellator, the use of LSH has been reduced in many institutions. The procedure-specific rating scale (CAT-LSH) can easily be modified by exchanging amputation of the cervix by circumcision and closing of the vagina. This change could make it relevant also for assessment of surgical competence during total laparoscopic hysterectomy.

CONCLUSIONS

The GOALS and CAT-LSH appear to have construct validity and high interrater reliability. Assessment of surgical competence during LSH is feasible in daily practice with objective rating scales like CAT-LSH and GOALS.

CONTRIBUTION TO AUTHORSHIP

The contribution of each author is as follows: J.M.G. contributed by designing and planning the study, collecting data, performing the statistical analyses and interpretation, writing and revising the manuscript, and final approval of the submitted manuscript. L.S. and E.F. contributed by planning the study, interpretation of statistical analysis, writing and critical revision of the manuscript, and final approval of the submitted manuscript. M.L. contributed by designing, facilitating and planning the study, collecting data, performing the statistical analyses and interpretation, writing and revising the manuscript, and final approval of the submitted manuscript.

DETAILS OF ETHICAL APPROVAL

The Regional Committee for Medical Research Ethics in eastern and southern Norway evaluated the study proposal and decided that ethical approval was not required.

ACKNOWLEDGMENTS

We are grateful to Frank W. Jansen, Thomas Lyons, and Anton Langebrekke for taking part in the expert group defining the surgical strategy for LSH and assisting us during the development of the CAT-LSH. Furthermore, we would like to thank Danilo Miskovic and his colleagues at Imperial College in London for inspiration, ideas and giving us permission to use CAT for surgical procedures as a model for the CAT-LSH. Finally, we appreciate the assistance by Anton Langebrekke and Jelena Kisic who evaluated all the video recordings of the LSH procedures included in the study.

REFERENCES

1. Sidhu RS, Grober ED, Musselman LJ, Reznick RK. Assessing competency in surgery: where to begin? *Surgery*. 2004;135(1):6-20.
2. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *Br Med J*. 2003;327(7422):1032-1037.
3. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Assessment of technical surgical skills. *Eur J Surg*. 2002;168(3):139-144.
4. Wulf G, Shea C, Lewthwaite R. Motor skill learning and performance: a review of influential factors. *Med Educ*. 2010;44(1):75-84.
5. Gjersvik P. Why is feedback so difficult? *J Norwegian Med Assoc*. 2014;134(6):591.
6. Losvik OK. Where the absence of criticism is the highest form of praise. *J Norwegian Med Assoc*. 2013;133(16):1690.
7. Shore M, Lefebrvre G, Grantcharov TP. Gynecology resident laparoscopy training:present and future. *Am J Obstet Gynecol*. 2015:288-301.
8. Ghaderi I. Technical skills assessment toolbox: a review using the unitary framework of validity. *Ann Surg*. 2015;261(2):215-262.
9. Aggarwal, Cutting Edges in Surgical Training: Via Media. 2013.
10. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.
11. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190(1):107-113.
12. Debas HT, Bass BL, Brennan MF, et al. American Surgical Association Blue Ribbon Committee Report on Surgical Education: 2004. *Ann Surg*. 2005;241(1):1-8.
13. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945-949.
14. Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. *J Am Coll Surg*. 2007;204(2):308-316.
15. Miskovic D, Ni M, Wyles SM, et al. Is competency assessment at the specialist level achiveable? *Ann Surg*. 2013;257:476-482.

16. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *Br J Obstet Gynecol*. 2008;115(7):908-916.

17. Chang L, Hogle NJ, Moore BB, et al. Reliable assessment of laparoscopic performance in the operating room using videotape analysis. *Surg Innov*. 2007;14(2):122-126.

18. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg*. 2011;202(4) 469.e6-80.e6.

19. Khan KS, Chien PF. Evaluation of a clinical test. I: assessment of reliability. *Br J Obstet Gynecol*. 2001;108 (6):562-567.

20. Oestergaard J, Larsen CR, Maagaard M, Grantcharov T, Ottesen B, Sorensen JL. Can both residents and chief physicians assess surgical skills? *Surg Endosc*. 2012;26(7):2054-2060.

21. Davis A, Mazmanian Paul, Fordis M, Harrison R. Accuracy of physician self-assessment compared with observed measures of competence. A systematic review. *J Am Med Assoc*. 2006;296(9):1094-1102.

II

European Journal of Obstetrics & Gynecology and Reproductive Biology: X 4 (2019) 100092

Contents lists available at ScienceDirect

European Journal of Obstetrics & Gynecology and Reproductive Biology: X

journal homepage: www.elsevier.com/locate/eurox

Development and validation of a general and easy assessable scoring system for laparoscopic skills using a virtual reality simulator

JM Goderstad[a,*], L Sandvik[b], E Fosse[c,e], M Lieng[d,e]

[a] Department of Surgery, Sørlandet Hospital, Sykehusveien, 4838, Arendal, Norway
[b] Oslo Center for biostatistics and epidemiology, Oslo University Hospital, Norway
[c] The Intervention Centre, Oslo University Hospital, Oslo, Norway
[d] Department of Gynecology, Oslo University Hospital, Oslo, Norway
[e] Institute of Clinical Medicine, University of Oslo, Oslo, Norway

ARTICLE INFO

Article history:
Received 25 November 2018
Received in revised form 3 July 2019
Accepted 4 August 2019
Available online 13 August 2019

Keywords:
Laparoscopy
Simulation
Virtual reality
Procedural training
Assessment of surgical training

ABSTRACT

Objectives: To develop and validate a scoring system for laparoscopic skills for five specific tasks on a virtual reality simulator.
Study design: A longitudinal, experimental, non-randomised study including 30 gynecologists and gynecological trainees at three hospitals. The participants were categorized as inexperienced (Group 1), moderately experienced (Group 2), and experienced (Group 3).
The study participants performed ten repetitions of three basic skill tasks, a salpingectomy and a laparoscopic supracervical hysterectomy on a virtual reality simulator. Assessment of skills was based on time, error parameters and economy of movements measured by the simulator. We used the results (mean and SD for each parameter in all tasks) of the four last repetitions performed by the experienced gynecologists as the basic for the scoring system. Performance equal to, and higher than, this mean score gave 2 points. A decrease of 1 SD from the mean gave 1 point. Every score below gave 0 points. The mean score for the inexperienced, moderately experienced and experienced study participants was compared.
Results: The mean scores in Task 1 were 3.4 (SD 0.6) in Group 1, 3.4 (SD 0.6) in Group 2 and 5.1 (SD 1.1) in Group 3, respectively. There was a statistically significant difference in score between Group 1 and 3 ($p = 0.01$), and group 2 and 3 ($p = 0.01$). In Task 2 no statistical significant differences were found. In Task 3, the total mean scores were 1.7 (SD 0.7) in Group 1, 1.9 (SD 0.9) in Group 2 and 2.8 (SD 0.5) in Group 3, respectively. The difference in score between study groups was statistically significant when comparing Group 1 and Group 3 ($p < 0.01$) and Group 2 and Group 3 ($p = 0.02$).
In Task 4, the difference in used time between group 1 and 3 was statistically significant ($p = 0.03$). In task 5 there was a significant difference in performance score between group 1 and 3 ($p = 0.01$).
Conclusions: There was significant difference in scores between the experienced and the inexperienced gynecologist in four out of five tasks.
The scoring system is easy assessable and can be used for summative and formative feedback in proficiency-based assessment.

1. Introduction

Surgical competence is a combination of surgical technique, experience and strategy. As the surgical technology evolves, along with a heightened awareness of patient safety and concerns, surgical guidelines and surgical skill training programs become critical. Surgical training results in improved surgical performance when it comes to operating time, efficiency and safety [1–6].

Surgical competence can be measured and evaluated using a variety of different validated scoring systems, such as GOALS (Global Operative Assessment of Laparoscopic Skills) and OSATS (Objective Structured Assessment of Technical Skills), among others. Simulators and pelvic trainers are recommended and accepted training tools, and permit practice of surgical skills in a safe environment without compromising patient safety [7]. Access to pelvic trainers and simulators alone may not be sufficient to ensure an effective skills training [8]. A training program should be

* Corresponding author.
E-mail address: jeanne.mette.goderstad@sshf.no (J. Goderstad).

http://dx.doi.org/10.1016/j.eurox.2019.100092
2590-1613/

based on proficiency and performance rather than a fixed number of repetitions or spent time [3]. Proficiency-based assessment is developed by experts performing the target procedure or a set of skills. The scores of the experts are then used to set the required score for passing the procedure or certification (summative feedback). Attaining basic psychomotor skills at an expert level on a surgical simulator system, could shorten the learning curves on real surgical procedures [1]. During development of surgical skills in the operating room, residents who have trained effectively on a surgical simulator are able to focus on the strategy of the procedure and the decision making during surgery, instead of focusing on the hand movements [9].

The objective of the study was to define and validate a scoring system for laparoscopic skills for five specific tasks on the simulator.

2. Materials and methods

This longitudinal, experimental, non-randomised study was conducted in accordance with the Declaration of Helsinki, and national and local regulations.

2.1. Inclusion

Gynecological trainees and consultants were invited to participate in the study. We recruited consecutively, until ten participants had been included in each study group. Prior to inclusion, all study participants received written information about the study, and they signed an informed consent for study participation. The surgical experience of each study participant at the time of inclusion was registered, and they were categorized into one of the following study groups:

Group 1: Inexperienced (performed less than 50 laparoscopic procedures, and previously never performed a laparoscopic hysterectomy)

Group 2: Intermediate experienced (previously performed more than 50 laparoscopic procedures, including more than five laparoscopic supracervical hysterectomies (LSH), but not performed total laparoscopic hysterectomies.

Group 3: Experienced (senior consultants performing total laparoscopic hysterectomy and surgery for women with deep infiltrating endometriosis).

2.2. Training

The training was carried out using the LAPmentor Express, Simbionix, 3D Systems, a portable, 2D non-haptic feedback simulator. At the first training session, all participants were given individual hands-on introduction to the simulator, and the tasks were presented. The program consisted of three basic skill tasks (Task 1, 2 and 3), a salpingectomy (Task 4) and a modified LSH (Task 5). All tasks were performed during each training session in a systematic order (Task 1–5, consecutively). This was repeated, dependent of available training time, up to maximum four times during one training session. The training was completed when all tasks had been performed ten times. The total training period was aimed to last between two and six weeks.

An instructor was present during all training sessions to assist the study participants in case they needed guidance on the simulator system or the tasks.

2.3. Description of the tasks

2.3.1. Task 1: two-handed maneuver

The task included exposure of nine balls embedded in jelly. A correctly exposed ball changed the color from red to green. All balls then had to be grabbed and placed into a basket.

This is a coordination task involving speed and precision. The objectives are to improve advanced bimanual skills, practice instrument manipulation and eye-hand coordination, and acquire tissue-handling skills.

The parameters measured were time (s), number of balls in the basket (n), total path length (cm) and instrument movement (number). In addition, number of errors was registered (only green balls should be grabbed).

2.3.2. Task 2: peg transfer

The participants lifted six objects from a pegboard with the left hand, transferred the object to the right hand, and placed them over the pegs on the pegboard. The process was then reversed.

The objectives are improved eye-hand coordination, use of both hands and depth perception.

The parameters measured were total time (s) and number of successfully moved objects (without loss and correctly placed on the pegboard) (n).

2.3.3. Task 3: pattern cutting

The participants used a grasper to apply traction exposing the best angle for the dominant hand to cut in the marked circle with accuracy.

The objective of this task is use of both hands and accuracy.

The parameters measured were total time (s) and errors (any deviation from the drawn line).

2.3.4. Task 4: left side salpingectomy

The participants used a grasper, scissors, and a bipolar forceps to remove the left tube. The total time used on the task (min) was registered. In case of an error (bleeding), it had to be corrected before commencing the salpingectomy.

2.3.5. Task 5: modified LSH

The participants were introduced to a step-by-step strategy starting on the left side and including [10].

1 Identification and division of the round ligament
2 Identification of the anterior leaf of the broad ligament and progressive cauterization of the ligament towards the middle medially paying attention to the bladder
3 Coagulation and division of the proper ovarian ligament and the fallopian tube
4 Division of the posterior leaf of the broad ligament
5 Identification, coagulation, and division of the uterine vessels
6 Step 1–5 was then performed at the right side
7 The cervix was exposed and the participant marked the correct level of amputation.

Total procedural time (min), total path length (cm), instrument movements (n) and errors (bleeding and improper respect of tissue/tissue handling) were registered. The registration started when the participant took hold of the left round ligament.

2.4. Statistics

All statistical tests were 2-sided, and $p \leq 0.05$ was considered statistically significant. Statistical analyses were performed using commercially available software (SPSS version 17.0; SPSS Inc., Chicago, IL). By using the findings in the publication by Fagerland et al, we found that the distribution of the variables for the parameters measured for each task (parameters of each task are described above) were sufficiently close to normal distribution for using the independent samples *t*-test [11].

2.5. Study sample size

The sample size calculation was based on the variable, “duration of task” from a procedural task (salpingectomy) in a study by Larsen CR et al. [9]. When calculating the sample size, we assumed that in the planned study, the difference in mean Total Time between the groups would be equal to the difference in median Total Time in the Larsen paper. The standard deviation of time in Larsens study is 90 s in the inexperienced group, and 40 s in the expert group. We assumed that similar standard deviation would be observed in our study. We furthermore assumed that the mean difference in time between Group 1 and Group 2 would be at least 90 s. It may be shown that if the true mean difference in time between these two study groups is at least 90 s, in a study with 80% test power and a significance level of 0.05, at least 10 physicians had to be included in each group. We consequently decided to include 10 study participants in each study group.

2.6. The scoring system

The results of the parameters of the four last repetitions of each task performed by the experienced participants were used as the base for the scoring system, and the mean and SD for these four repetitions was registered for each parameter in each task. In each repetition, a performance equal to, or higher than, the mean of the experienced participants in a registered parameter gave a score of two points. Up to one SD decrease from the mean in each parameter resulted in a score of one point. Every score below one SD gave 0 points. The scores from each parameter in each task (in Task 1, 2, 3, and 5) were added to give the total task score. Since the different tasks have different number of registered parameters, the maximum score differed between the different tasks (from 4 to 10), as illustrated in Figs. 1–3 and 5. In Task 4, we used time (min), as this was the only parameter in this task.

3. Results

3.1. Study participants

The mean age of the study participants in the three study groups was 31 years (SD 5.0) in Group 1, 36 years (SD 4.9) in Group 2 and 51 years (SD 7.3) in Group 3. All included study participants were right-handed. None of the study participants had any previous experience with the LapMentor simulator.

3.2. Training sessions

The study took place from September 2013 until May 2014. Some of the planned training sessions had to be postponed because of competing clinical activities and unexpected responsibilities. However, all study participants completed the training. The median total training period was 48 days (range 14–63 days) in

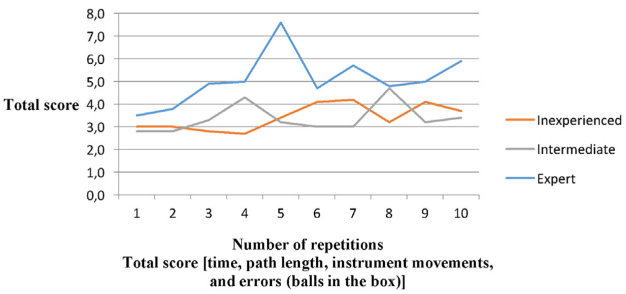


Fig. 1. Task 1 (Two-handed maneuver).

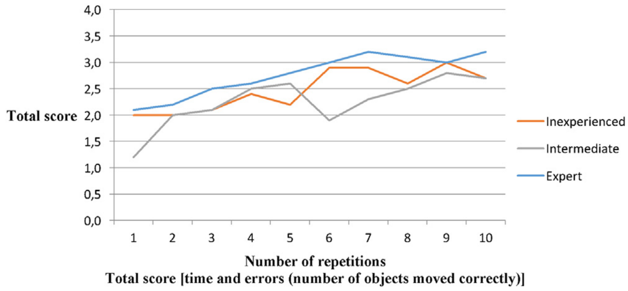


Fig. 2. Task 2 (Peg transfer).

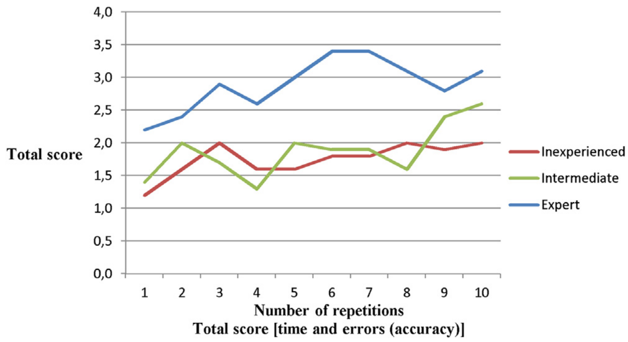


Fig. 3. Task 3 (Pattern cutting).

Group 1, 19 days (range 7–61 days) in Group 2 and 25 days in Group 3 (range 4–60 days), respectively.

3.3. Scores of performance

The performance scores of all five tasks are presented in Figs. 1–5.

The mean scores in Task 1 were 3.4 (SD 0.6) in Group 1, 3.4 (SD 0.6) in Group 2 and 5.1 (SD 1.1) in Group 3, respectively. There was a statistically significant difference in score between Group 1 and 3 ($p = 0.01$), and group 2 and 3 ($p = 0.01$). The difference in score between Group 1 and Group 2 was not statistical significant ($p = 0.85$).

The total mean scores in Task 2 were 2.5 (SD 0.7) in Group 1, 2.3 (SD 0.7) in Group 2 and 2.8 (SD 0.3) in Group 3, respectively. In Task 2, no statistical significant differences in total score between the study groups were found (Group 1 vs. Group 3, $p = 0.1$, Group 1 vs. Group 2, $p = 0.5$ and Group 2 versus Group 3, $p = 0.1$).

The total mean scores in Task 3 were 1.7 (SD 0.7) in Group 1, 1.9 (SD 0.9) in Group 2 and 2.8 (SD 0.5) in Group 3, respectively. The

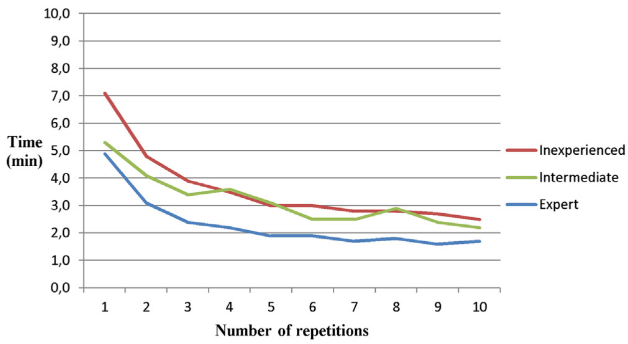


Fig. 4. Task 4 (Salpingectomy).

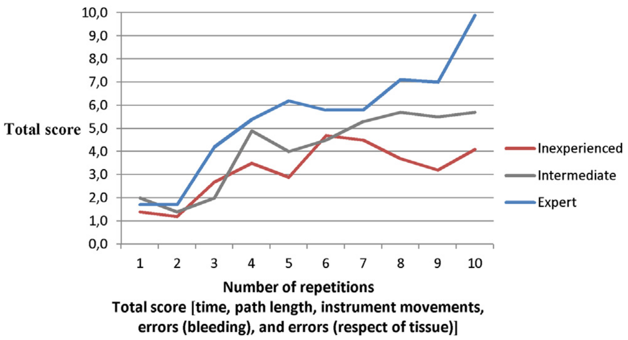


Fig. 5. Task 5 (Modified laparoscopic supracervical hysterectomy).

difference in score between study groups was statistically significant when comparing Group 1 and Group 3 ($p < 0.01$) and Group 2 and Group 3 ($p = 0.02$). There was no statistical significant difference in mean score when comparing Group 1 and Group 2 ($p = 0.60$).

The mean time used in Task 4 was 3.6 min (SD 1.4 min) in Group 1, 3.2 min (SD 0.9 min) in Group 2 and 2.3 min (SD 1.0 min) in Group 3, respectively. The difference in used time between group 1 and 3 was statistically significant ($p = 0.03$). There was no statistical significant difference in time when comparing Group 1 and Group 2 ($p = 0.45$) and Group 2 and Group 3 ($p = 0.06$).

The total mean performance score in Task 5 was 3.2 (SD 1.5) in Group 1, 4.0 (SD 1.6) in Group 2 and 5.3 (SD 1.8) in Group 3, respectively. There was a significant difference in performance score between group 1 and 3 ($p = 0.01$). The difference in mean score when comparing Group 1 and 2, and Group 2 and 3 was not statistically significant $p = 0.24$ and $p = 0.1$, respectively.

4. Comment

The results of this study showed a statistically significant difference in mean score when comparing the performance of experienced and inexperienced gynecologists in four out of five tasks in a standardized training program. Hence, the scoring system has validity for assessment of performance on the simulator.

The participants in Group 1 had some laparoscopic experience prior to inclusion. Inclusion of students in Group 1would probably resulted in larger differences between the groups. However, as the objective was to validate the scoring system for use in a clinical setting, we chose to include registrars that had started their laparoscopic training.

The participants in Group 2 were heterogeneous in respect to surgical experience. This might explain lack of significant differences between the groups in some tasks. The results of Group 2 are less relevant as the clinical importance of the scoring system is to differentiate between Group 1 and 3. Consequently, only comparing Group 1 and 3 probably would have improved the validation of the scoring system without reducing the quality of the study.

The lack of significant differences between groups in different tasks could furthermore be related to the level of difficulty of the tasks. The effect of training, measured as increase in total score, furthermore varied between the different tasks. This might be explained by the true value of the tasks. Given the results of group 1 and 2 in task 1, the value of this particular task can be questioned. The study participants were categorized into the study groups based on number and types of previous performed laparoscopic procedures. Previous authors have argued that previous performed procedures do not necessarily represent actual clinical competence [12–14]. Consequently, a different selection of study participants into the different groups might have influenced the mean performance and consequently the difference between the study groups. Learning curves express the relationship between an outcome variable, a score, and the number of repetitions of a given task, and can be used to determine when additional training is less likely to increase performance as well as to individualize training programs. Previous studies have investigated the influence on training schedules on surgical technical skills and show superiority for distributed training [15,16]. However, the interval between training sessions in a distributed training model may affect the outcome of the training. If the interval between the training sessions is too long, the retention of skills is influenced [16]. This may have been the case for some of the study participants especially in group 1. For practical reasons, ten repetitions of tasks were performed in our study setting. Brunner et al. have reviewed the literature describing the optimal number of repetitions during training [17]. Their data demonstrated that an initial plateau is reached after eight repetitions, but that the overall best score result was reached after 21–29 repetitions. Consequently, more repetitions in this study might have affected the training outcomes in all study groups. One of the strengths of virtual reality simulators is the standardized setting. This makes the setting fair for the participants, and can motivate them to participate in the study and complete the training period. This might also have contributed to all participants completing the training in this study. The results of this study demonstrate that different parameters in training tasks can be combined and integrated in a total score, which enables summative and formative feedback. The same principle can be done with GOALS, OSATS and other systems, but with these assessment tools an observer is mandatory to perform the assessment. Once the proficiency level of an exercise is set, it often will serve as a motivation factor for trainees wanting to increase their surgical skills.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank all the gynecologists who participated in the study, and their colleagues who covered their duties in the clinic while the training was performed.

References

[1] Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ 2009;338:b1802.
[2] Aggarwal R, Tully A, Grantcharov T, Larsen CR, Miskry T, Farthing A, et al. Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy. BJOG Int J Obstet Gynaecol 2006;113(12):1382–7.
[3] Strandbygaard J, Bjerrum F, Maagaard M, Rifbjerg Larsen C, Ottesen B, et al. A structured four-step curriculum in basic laparoscopy: development and validation. Acta Obstet Gynecol Scand 2014;93(4):359–66.
[4] Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann surg 2002;236(4)458–63 Discussion 63-4.
[5] Kundhal PS, Grantcharov TP. Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. Surg Endosc 2009;23(3):645–9.
[6] Gauger PG, Hauge LS, Andreatta PB, Hamstra SJ, Hillard ML, Arble EP, et al. Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. Am J Surg 2010;199(1):72–80.
[7] Aggarwal R, Ward J, Balasundaram I, Sains P, Athanasiou T, Darzi A. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. Ann Surg 2007;246(5):771–9.

[8] van Dongen KW, van der Wal WA, Rinkes IH, Schijven MP, et al. Virtual reality training for endoscopic surgery: voluntary or obligatory? Surg Endosc 2008;22(3):664–7.
[9] Larsen CR, Grantcharov T, Aggarwal R, Tully A, Sorensen JL, Dalsgaard T, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc 2006;20(9):1460–6.
[10] Goderstad JM, Sandvik L, Fosse E, Lieng M. Assessment of Surgical Competence: Development and Validation of Rating Scales Used for Laparoscopic Supracervical Hysterectomy. J Surg Educ 2016;73(4):600–8.
[11] Fagerland MW, Sandvik L. Performance of five two-sample location tests for skewed distributions with unequal variances. Contemp Clin Trials 2009;30(5):490–6.
[12] Bjerrum F, Strandbygaard J, Rosthoj S, Grantcharov T, Ottesen B, Sorensen JL. Evaluation of procedural simulation as a training and assessment tool in general surgery-simulating a laparoscopic appendectomy. J Surg Educ 2017;74(2):243–50.
[13] Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. Adv health Sci Educ: Theory Pract 2015;20(3):829–34.
[14] Korndorffer Jr. JR, Kasten SJ, Downing SM. A call for the utilization of consensus standards in the surgical education literature. Am J Surg 2010;199(1):99–104.
[15] Verdaasdonk EG, Stassen LP, van Wijk RP, Dankelman J. The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. Surg Endosc 2007;21(2):214–9.
[16] van Dongen KW. Distributed versus massed training:efficiency of training psychomotor skills. Surg Tech Dev 2011;1:40–2.
[17] Brunner WC, Korndorffer Jr. JR, Sierra R, Massarweh NN, Dunne JB, Yau CL, et al. Laparoscopic virtual reality training: are 30 repetitions enough? J Surg Res 2004;122(2):150–6.

III

Research article

Development and validation of a curriculum for laparoscopic supracervical hysterectomy

J.M. Goderstad[1], E. Fosse[2,3], L. Sandvik[4], M. Lieng[3,5]

[1]Department of Surgery, Sørlandet Hospital, Sykehusveien 1, 4838 Arendal, Norway; [2]The Intervention Centre, Oslo University Hospital, 0424 Oslo, Norway; [3]Institute of Clinical Medicine, University of Oslo, 0316 Oslo, Norway; [4]Center for biostatistics and epidemiology, Oslo University Hospital, 0373 Oslo, Norway; [5]Division of Obstetrics and Gynecology, Oslo University Hospital, 0424 Oslo, Norway.

Correspondence at: Marit Lieng,, Division of Obstetrics and Gynecology, Oslo University Hospital, Kirkeveien 166, 0450 Oslo, Norway, email: m.lieng@online.no

Abstract

Study Objective: **To develop and validate a three-step curriculum for laparoscopic supracervical hysterectomy (LSH) designed for a busy clinical setting.**

Methods: **Single-centre, prospective, cohort study. Twelve eligible gynaecological trainees were included (group 1). The theoretical part (step 1) was a validated multiple-choice test. The practical part (step 2) consisted of five tasks on a virtual reality simulator. The participants had to reach a pre-defined proficiency level before advancing to performing a LSH (step 3). The validation of the curriculum was based on the surgical performance. The surgical procedure was recorded and assessed by two experts using Global Operative Assessment of Laparoscopic Skills (GOALS) and Competence Assessment Tool – Laparoscopic Supracervical Hysterectomy (CAT-LSH). The scores were compared with scores from gynaecological trainees who performed their first LSH without virtual reality simulator training (group 2).**

Results: **Ten trainees completed the curriculum and performed a LSH that was recorded and evaluated. Mean duration of the training period (step 1 and 2) was 57 days (SD 26.0), and mean training time spent on the simulator to reach the pre-set proficiency level was 173 min (SD 49). The mean GOALS score was 18.5 (SD 5.8) in group 1 and 13.6 (SD 3.3) in group 2, p=0.027. The mean CAT-LSH score of the performance of the hysterectomy was 42.1 (SD 6.9) in group 1 and 34.8 (SD 4.3) in group 2, p= 0.009.**

Conclusions: **Trainees who completed the curriculum appeared to have a higher performance score compared with trainees who did not perform structured training.**

Keywords: **Laparoscopy, simulation training, proficiency-based training, certification, curriculum.**

Introduction

It is recommended that surgical residents undergo initial laparoscopic skills training outside the operating room (Reznick et al., 2006; Seymour et al., 2002; Larsen et al., 2009; Campo et al., 2014; Gala et al., 2013; Sroka et al., 2010). However, in spite of convincing research reports demonstrating the advantages of structured laparoscopic skills training, the implementation of structured training curricula remains challenging. This might be explained by limiting factors such as logistics, equipment, clinical tasks and working hours.

Along with the increased implementation of minimally invasive surgery for common surgical procedures, there has been a concomitant reduction of participation of junior-level residents. The trend of less surgical procedures among residents has significant implications for surgical resident education (Mullen et al., 2016). It is consequently essential that a part of the surgical training takes part outside the operating theatre. Otherwise, the practical skills for future surgeons might be negatively influenced.

For a surgical training curriculum to be successful, several elements are required. It is essential that the curriculum contains a cognitive component, a practical component, and subsequent supervised training in the actual clinical setting (Strandbygaard et al., 2014). A successful surgical training curriculum also depends on trainee, faculty and employer commitment (Stefanidis et al., 2009). Scoring systems and proficiency-based training can give summative and formative feedback that motivates the trainees. Logistics that facilitates distributed training with defined training hours is essential. Hence, the need for structured training in a busy clinical setting must be acknowledged by the employer in order to implement a structured training programme successfully. Education of health personnel, including registrars, is one of the major tasks of teaching hospitals. However, in our experience, structured training for registrars is often not prioritised in clinical departments. The objective of the study was to develop and validate a curriculum for laparoscopic supracervical hysterectomy (LSH) that increased the trainees' surgical performance and was designed to be feasible to completed in a busy clinical setting.

Methods

The study was a single-centre, cohort study performed at a Norwegian university hospital. All junior trainees at the department of gynaecology were invited to participate. To be eligible for study participation, they had to be able to perform basic laparoscopic procedures supervised by a consultant, and about to move on to perform more complex procedures like hysterectomy. Registrars eligible for study inclusion should not have performed more than 50 laparoscopic procedures previously, and they should plan to continue employment at the department for the next six months. Prior to inclusion, all study participants received written information about the study, and they signed an informed consent for participation. The study participants were followed up according to the study flowchart (Figure 1).

The curriculum consisted of three steps. All study participants first underwent a theoretical knowledge multiple-choice test within basic laparoscopy (step 1). We used questions from a test previous published by Strandbygaard et al., 2013. As the curriculum was designed for laparoscopic supracervical hysterectomy (LSH), we added six questions related to this particular procedure. These questions were evaluated by three international experts in gynaecological laparoscopy before they were included in the theoretical test. The aim of the theoretical test was to stimulate the study participants to obtain theoretical knowledge related to laparoscopy in general, as well as the specific procedure before they started the training and performed the first surgical procedure. Wrong answers in the test did not have any consequences for the registrars, but all wrong answers were discussed to make the registrar able to give the correct answer.

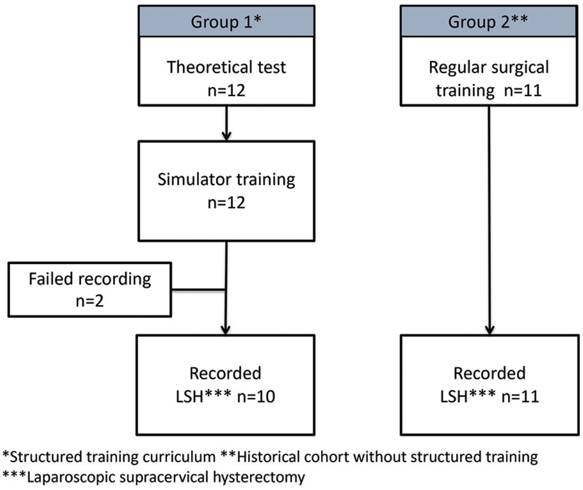


Figure 1: Study flow chart.

The study participants then underwent a structured individual laparoscopic training programme (step 2). The training was carried out using the Simbionix, LAPmentor Express, 3D, VR simulator. At the first training session, all study participants were given a standardised hands-on introduction to the Simbionix LAPmentor system, and an oral presentation as well as a video presentation of the different tasks included in the training programme. The tasks had varying complexity and consisted of three basic skill tasks, a salpingectomy and a LSH. We aimed for distributed training, meaning short training periods, with rest periods in between. We planned three repetitions of each task at each training session and estimated that the participants needed at least three sessions to reach the pre-set proficiency level.

Description of the tasks included in step 2:

Task 1: Two-handed manoeuver

The task included exposure of nine balls embedded in jelly. A correctly exposed ball changed the colour from red to green. All balls then had to be grabbed and placed into a basket (Figure 2).

This is a coordination task involving speed and precision. The objectives are to improve advanced bimanual skills, practice instrument manipulation

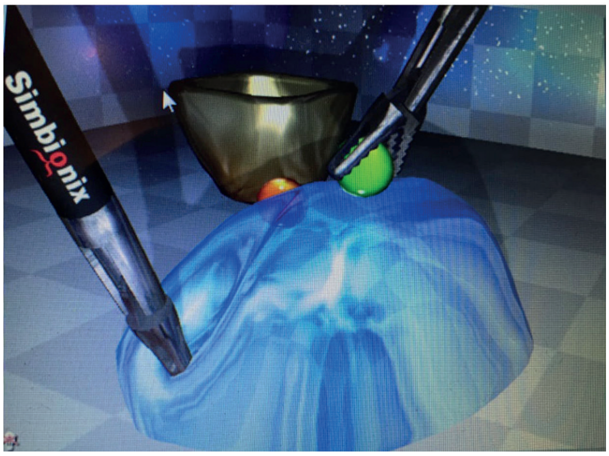


Figure 2: Task 1. Two-handed manoeuver.

and eye-hand coordination, and acquire tissue-handling skills.

The parameters measured were time (s), number of balls in the basket (n), total path length (cm) and instrument movement (number). In addition, number of errors was registered (only green balls should be grabbed).

Task 2: Peg transfer

The participants lifted six objects from a pegboard with the left hand, transferred the object to the right hand, and placed them over the pegs on the pegboard. The process was then reversed (Figure 3).

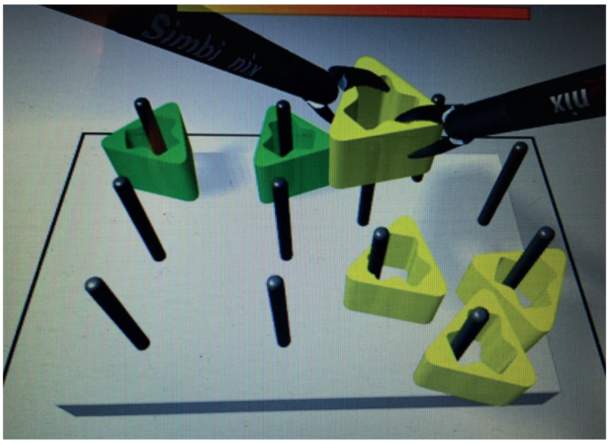


Figure 3: Task 2. Peg transfer.

The objectives are improved eye-hand coordination, use of both hands and depth perception, and the measured parameters were total time (s) and number of successfully moved objects (without loss and correctly placed on the pegboard) (n).

Task 3: Pattern cutting

The participants used a grasper to apply traction exposing the best angle for the dominant hand to cut in the marked circle with accuracy. The objective of this task is use of both hands and accuracy. The parameters measured were total time (s) and errors (any deviation from the drawn line).

Task 4: Left side salpingectomy

The participants used a grasper, scissors, and a bipolar forceps to remove the left tube. The total time used on the task (min) was registered. In case of an error (bleeding), it had to be corrected before commencing the salpingectomy.

Task 5: Modified LSH

The participants were introduced to a step-by-step strategy of the procedure starting on the left side and including: identification and division of the round ligament, identification of the anterior leaf of the broad ligament and progressive cauterisation of the ligament towards the middle medially paying attention to the bladder, coagulation and division of the proper ovarian ligament and the fallopian tube, division of the posterior leaf of the broad ligament and identification, coagulation, and division of the uterine vessels (Figure 4). The same steps were then performed at the right side. Finally, the cervix was exposed and the participant marked the correct level for amputation. In this task, total procedural time (min), total path length (cm), instrument movements (n) and errors (bleeding and improper respect of tissue/tissue handling) were registered. The registration started when the participant took

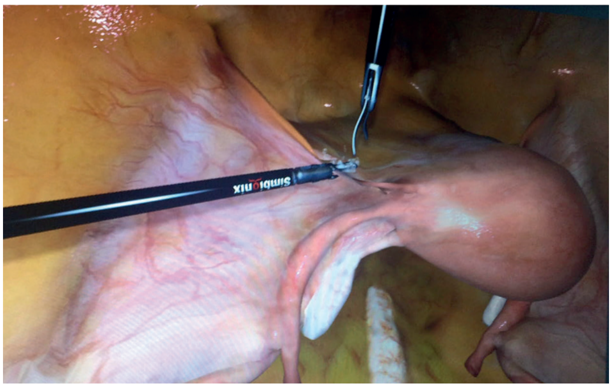


Figure 4: Task 5. Laparoscopic supracervical hysterectomy.

hold of the round ligament on the left side.

The registered parameters in each task were combined, giving a total score of performance for evaluation of skills of each study participant in each task.

The study participants performed the tasks independently. An instructor (JMG) was present during all training sessions in order to give feedback, assist the study participants in case they needed guidance on the simulator system, or information regarding the different tasks.

Table I. – Global Operative Assessment of Laparoscopic Skills (GOALS).

Domains	1	2	3	4	5
Depth perception	Constantly overshooting target, hits backstop, wide swings, slow to correct.		Some overshooting or missing plane but corrects quickly.		Accurately directs instruments in correct plane to target.
Bimanual dexterity	Use of one hand, ignoring non-dominant hand, poor coordination between hands.		Use of both hands but does not optimize interactions between hands to facilitate conduct of operation.		Expertly uses both hands in a complementary manner to provide optimal working exposure.
Efficiency	Uncertain, much wasted effort, many tentative motions, constantly changing focus of operation, or persisting at a task without progress.		Slow, but planned and reasonably organized.		Confident, efficient and safe conduct of operation, maintaining focus on component of procedure until better done approach.
Tissue handling	Rough, tears tissue by excessive traction, injures adjacent structures, poor control of coagulation device, grasper frequently slips off.		Handles tissues reasonably well with some minor trauma to adjacent tissues, eg coagulation of non- target tissue, occasional slipping of grasper.		Handles tissue very well with appropriate traction on tissues and negligible injury of adjacent structures. Uses energy sources appropriately but not excessively.
Autonomy	Unable to complete entire procedure, even in a straight forward case and with extensive verbal guidance.		Able to complete operation safely with moderate prompting.		Able to complete operation independently without prompting.
Level of difficulty	Easy exploration and dissection.		Moderate difficulty (eg, mild inflammation, scarring, adhesions, obesity, severity of Disease.		Extremely difficult (eg, severe inflammation, scarring, adhesion, obesity, or severity of disease).

The total training period (step 2) was aimed to last between two and six weeks. The different tasks were repeated until the participant reached the pre-set level of proficiency. This level was defined by results from a previous study including experienced laparoscopic surgeons performing the same tasks (Goderstad et al., 2019).

When the study participants reached the required proficiency level, they performed a LSH assisted by an experienced surgeon (step 3). The procedure was recorded and evaluated by two blinded experienced surgeons, using two validated scoring systems, Global Operative Assessment of Laparoscopic Skills (GOALS) and Competence Assessment Tool - Laparoscopic Supracervical Hysterectomy (CAT-LSH) (Goderstad et al., 2016). In GOALS, the performance within six domains (depth perception, bimanual dexterity, efficiency, tissue handling, autonomy and level of difficulty) are evaluated and given a score from 1 – 5 (Table I). The scores of each domain are added to a total score. Hence, the lowest possible GOALS score is 6, and the highest 30.

Evaluation using CAT-LSH includes the scoring of four procedure specific surgical variables within four different steps of the LSH procedure (Table II). All variables in each step are given a score from 1-4, and then added to a total score. The lowest possible CAT-LSH score is consequently 16, and the highest 64. The results were then compared with a cohort of trainees at similar surgical competence levels who performed their first LSH assisted by an experienced surgeon without preforming any preoperative training on a simulator (group 2) (Goderstad et al., 2016). The inclusion criteria for this group were identical as for the training group (group 1); the surgical procedure was performed in the same standardised manner and by using similar instruments, and the same two blinded experienced surgeons evaluated the recorded procedures using the two same validated scoring systems (GOALS and CAT-LSH).

Sample size calculation: The sample size calculation was based on the total GOALS score. In a previous simulation study with eight inexperienced and 13 experienced surgeons, mean difference in GOALS between the groups was 5.8, and the standard deviation in each group was 4.0 (Larsen et al., 2012). We assumed that we would obtain equivalent GOALS scores in our study, and planned to use independent samples t-test with 5% significance level when comparing the groups. It may then be shown that in order to achieve 80 % test power, at least 10 trainees had to be included in each study group. Assuming a 20 % drop-out rate, we decided to include 12 trainees in the intervention group (group 1). Notably, we already had access to data from 11 inexperienced trainees (group 2) from a previous study (Goderstad et al., 2016).

Statistical analyses: The data was analyzed using SPSS 23.0 (SPSS Inc., Chicago, IL). All statistical tests were performed two sided at the significance level of 5 %. A two-sided Independent t-test was used when comparing normally distributed continuous variables.

Ethical approval: The study was conducted in accordance with the Declaration of Helsinki and national and local regulations. The Regional Committee for Medical Research Ethics in eastern and southern Norway and the personal data officer at Oslo University Hospital approved the study protocol.

Results

The study took place from March 2015 to August 2016. All 12 study participants completed the theoretical test and the structured simulation training and reached the required proficiency level. During the surgical procedure (step 3), recording of the procedure failed in two cases, leaving 10 recorded procedures available for blinded evaluation (Figure 1). The mean duration of the training period (step 1-2) was 57.0 days, (SD 26.0). Mean time spent on the different tasks was 12.5 min (SD 3.7), 17.8 min (SD 7.7), 22.7 min (SD 13.5), 22.9 min (SD 10.1) and 97.8 min (SD 36.9) for task 1, 2, 3, 4 and 5, respectively. Mean total duration for all five tasks was 173.0 min (SD 49.0) (Figure 5). The average number of repetitions to reach proficiency level for all tasks was 42.0 (SD 8.0) Mean number of repetitions was 8.8 (SD 3.0), 7.8 (SD 3.0), 9.6 (SD 4.8), 6.5 (SD 2.7), and 9.1 (SD 3.0), for task 1, 2, 3, 4 and 5, respectively (Figure 6).

In the validation of the curriculum, the mean scores following the assessment by the two experts were 18.5 (SD 5.8) for GOALS, and 42.1 (SD 6.9) for CAT-LSH, respectively. The scores in the group of trainees who did not undergo the structured training (group 2) was 13.6 (SD 3.3) for GOALS and 34.8 (SD 4.3) for CAT-LSH.

The differences in GOALS and CAT-LSH mean scores between group 1 and 2 were statistically significant, p=0.027 for GOALS and p= 0.009 for CAT-LSH.

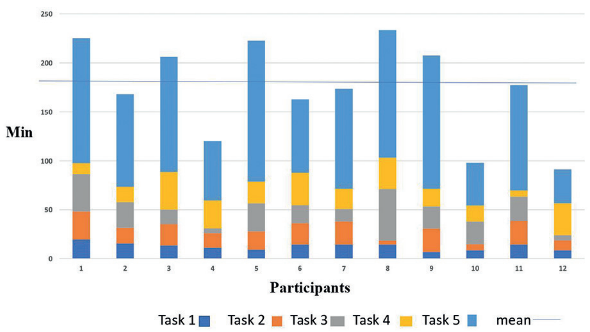


Figure 5: Time spent to reach proficiency level for each task.

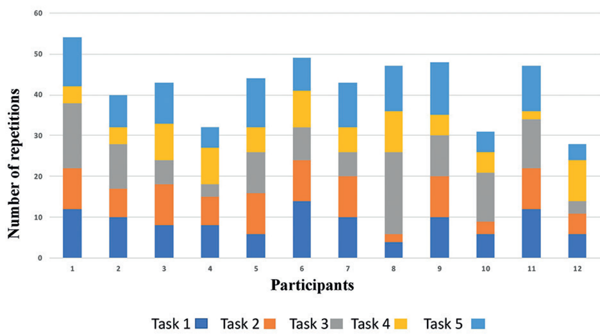


Figure 6: Number of repetitions needed to reach proficiency level for each task.

Discussion

The training curriculum appeared to have a positive effect on the surgical performance, as trainees who completed the curriculum had a higher performance score on their first laparoscopic hysterectomy compared to trainees without structured training. Our results are supported by findings of previous studies, evaluating the surgical performance following structured training outside the operating theatre (Larsen et al., 2012). By using the mean of the experts' performance it is possible to set a proficiency level for any training procedure on a box, a simulator or tasks in a dry or wet lab (Goderstad et al., 2019). Training in order to reach a defined proficiency level consequently appears to improve the surgical performance in the operating room. This finding is of importance to clinical practice. In 2014, six leading international societies within gynaecology recommended that each hospital teaching endoscopic surgery should make an endoscopic dry lab for training available in order to improve the proficiency of the endoscopic surgery skills of the physician. In Norway, all hospitals have at least one box trainer and some also have virtual reality simulators. This means that the physical tools for skills training are available. One of the advantages of simulators is immediate objective feedback on different

Table II. – Competence Assessment Tool-Laparoscopic Supracervical Hysterectomy (CAT-LSH) (1/2).

Tasks	Instruments	Tissue	Errors	End-product
Lig. mobilisation. Adequate exposure and dividing of the round ligament and the broad ligament.	Use of instruments: Uncoorninated – Stiff and uncontrolled movements,overshooting Hesitant – Controlled movements but hesitant and inefficient Skillfull – Smooth, controlled and meaningfull movements Versatile – Masterfull instrument use, effective movements n/a	Use of non-dominant hand (NDH): Sagnant – NDH does not move Lagging – NDH is adjusting with delay or without efficency Meaningfull – Meaningfull adjustment of NDH to improve exposure Forward looking – Strategic and intelligent adjustment by NDH n/a	This task was performed with: Complication – Bleeding Near miss – To close to the pelvic wall. Bloody dissection. No damage – No damage Tissue protective – Performed with best possible tissue protection n/a	Was the ligaments divided safely? No – Divided in an unfavorable distance from the uterus. Vaguely – No access to the vesicouterine space Yes – Safe divition. Anatomically – Crystal clear demonstration of anatomy n/a
Adnexa The proper ovarian ligament and fallopian tube are coagulated and divided	Use of bipolar forceps, scissors or harmonic: Uncoorni-nated – Stiff and uncontrolled movements, overshooting Hesitant – Controlled movements but hesitant and inefficient Skillfull – Smooth, controlled and meaningfull movements Versatile – Masterfull instrument use, effective movements n/a	Use of non-dominant hand: Sagnant – NDH does not move Lagging – NDH is adjusting with delay or without efficency Meaningfull – Adjustment of NDH to improve exposure Forward looking – Strategic and adjustment by NDH n/a	This task was performed with: Complication – Bleeding Near miss – To close to the ovarian tissue/uterus No damage – No damage to the ovaries. Tissue protective – Performed with best possible tissue protection. n/a	Was the proper ovarian ligament and the fallopian tube coagulated and divided correctly? No – Divided in an unfavorable distance from the uterus. Vaguely – To close to the ovarian tissue. Yes – Main structures identified and divided correctly. Anatomically – Crystal clear demonstration of anatomy n/a
Vascular controle Identify the uterine vessels, coagulate and divide them	Use of bipolar forceps and scissors: Hazardios – Uncontrolled movements Laborious – Awkward and repeated unnecessary attempts Efficent – Instruments accurately placed and engaged Masterly – Highly efficient and safe use of instruments n/a	Dissection of the uterine vessels: Hazardios – Insufficient view, uncontrolled movements Laborious – Repeated unnecessary attempts Efficent – Instruments accurately placed and engaged Masterly – Highly efficient and safe use of instruments n/a	This task was performed with: Complication – Bleeding, the vessels was not identified Near miss – Several attempts at several places to secure the vessels Efficent – Visualisation of the vessels Ideal precision – Smooth and efficient dissection n/a	Are the vessels identified and secured at the right level ? Uncontrolled – Vessels not secured Imprecise – Vessels not accurately secured Safe – Vessels secured before divition Flawless – Perfectly secured before divition n/a

performance parameters such as time, total path length, instrument movements and errors (bleeding and improper respect of tissue/tissue handling), and available learning curves.

We chose supracervical hysterectomy as the surgical procedure in the presented curriculum. The principle used in this study with a pre-set level of competence as a goal for the training, may be used for any surgical procedure. Departments that implement competence-based education for registrars, will after some time get experience with the average time needed for trainees to reach competence. This facilitates adding surgical training as a part of the registrars' regular tasks. Consequently, the schedule of the trainees' daily work can be organized accordingly. When a structured training curricula is implemented within a department, the trainees, the trainers and the employers will expect that the trainees practice, and that they have to reach a preset level of competence before they move on to surgical procedures in the operating room.

Another implication of introducing a curriculum with goal-oriented training, in contrast to time-based training, is the time spent on the simulator. When the trainees have reached the defined level of performance, they must move on to task of increased difficulty to further develop their skills (Strandbygaard et al., 2014; Guadagnoli and Lee, 2004). This knowledge should be used to make the working hours and surgical competence development of the residents as efficient as possible. When the registrars have reached the required competence level on the simulator, they move on to more advanced tasks or surgery, as spending time repeating procedures that are mastered on the simulator or on the box trainer have limited effect on further skills development.

The retention of skills is affected with time (Verdaasdonk et al., 2007; Guadagnoli and Lee, 2004). The retention of skills is a factor that must be taken into consideration for employers who choose to invest in training tools and implementation of a curriculum. It is essential to let the trainees move on to the operating room when they have finished the curriculum. In our opinion, this would furthermore improve the motivation of the trainees, as there is a clear and pre-defined goal of the training.

An education programme where you get a date for your first procedure when finishing the curriculum is an ideal situation. This might be a challenge to employers due to logistics. However, it might commit the trainees to implement the curriculum when they know the employer's expectations and the positive consequences of the opportunity to do surgery.

The relative high variation of the duration of the training period in the study is related to individual variation of time to reach the proficiency level, but also working hours and logistics. We aimed for distributed training with sessions every week, since this is known to give the best results when developing psychomotor endoscopic skills (Van Dongen, 2011). The trainees performed their ordinary clinical duties on the ward during the training period. This led to an extension of training sessions for some of the participants.

The variation of the length of the training period might be a limitation of the study, but is also a strength because it increases the validity. It was performed in a busy clinical department without adjustment of clinical activity to facilitate training. In our opinion, this improves the quality and feasibility of this curriculum. Another strength of the study is that the surgical performance after training was assessed using validated assessment tools, and we could compare the performance with a comparable group of trainees that had not undergone systematic training.

Our results indicate that the accomplishment of a proficiency level on the simulator, predict that the trainees will perform surgery at a more automated level than trainees without training. Consequently, implementation of proficiency-based training in a dry lab before training in the operating room, might contribute to make our residents good enough for our patients (Campo et al., 2012). Use of validated scoring systems during surgery will further add value both to registrars and their supervisors in order to monitor surgical skills development and identify potential need for further training.

In conclusion, trainees who implemented the curriculum appeared to have a higher performance score on their first laparoscopic hysterectomy compared to trainees who did not perform structured training. By using the mean of the experienced surgeons' performance, a proficiency level for any training procedure in order to enhance surgical skills and patient safety may be set.

Acknowledgments : Thanks to Jelena Kisic og Anton Langebrekke who have made a valuable contribution by assessing and scoring the video recorded LSH procedures.

Funding information: Oslo University Hospital funded the study (salary for first author).

References

Campo R, Puga M, Meier Furst R et al. Excellence needs training "Certified programme in endoscopic surgery". Facts Views Vis Obgyn. 2014;6:240-4.

Campo R, Molinas CR, De Wilde RL et al. Are you good enough for your patients? The European certification model in laparoscopic surgery. Facts Views Vis Obgyn. 2012;4:95-101.

Gala R, Orejuela F, Gerten K et al. Effect of validated skills simulation on operating room performance in obstetrics and gynecology residents: a randomized controlled trial. Obstet Gynecol. 2013;121:578-84.

Goderstad JM, Sandvik L, Fosse E et al. Assessment of surgical competence: development and validation of rating scales used for laparoscopic supracervical hysterectomy. J Surg Educ. 2016;73:600-8.

Goderstad JM, Sandvik L, Fosse E et al. Development and validation of a general and easy assessable scoring system for laparoscopic skills using a virtual reality simulator. Eur J Obstet Gynecol Reprod Biol X. 2019;4:100092.

Guadagnoli MA, Lee TD. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. J Mot Behav. 2004;36:212-24.

Larsen CR, Soerensen JL, Grantcharov TP et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ. 2009;338:b1802.

Larsen CR, Oestergaard J, Ottesen BS et al. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta Obstet Gynecol Scand. 2012;91:1015-28.

Mullen MG, Salerno EP, Michaels AD et al. Declining operative experience for junior-level residents: is this an unintended consequence of minimally invasive surgery? J Surg Educ. 2016;73:609-15.

Reznick RK, MacRae H. Teaching surgical skills-changes in the wind. New Eng J Med. 2006;355:2664-9.

Seymour NE, Gallagher AG, Roman SA et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg. 2002;236:458-64.

Sroka G, Feldman LS, Vassiliou MC et al. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. Am J Surg. 2010;199:115-20.

Stefanidis D, Heniford BT. The formula for a successful laparoscopic skills curriculum. Arch Surg. 2009;144:77-82.

Strandbygaard J, Maagaard M, Larsen CR et al. Development and validation of a theoretical test in basic laparoscopy. Surg Endosc. 2013;27:1353-9.

Strandbygaard J, Bjerrum F, Maagaard M et al. A structured four-step curriculum in basic laparoscopy: development and validation. Acta Obstet Gynecol Scand. 2014;93:359-66.

Van Dongen KW. Distributed versus massed training:efficiency of training psychomotor skills. Surgical Techniques Development. 2011;1:40-2.

Verdaasdonk EG, Stassen LP, van Wijk RP, Dankelman J. The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. Surg Endosc. 2007;21:214-9.